

Solución a la práctica 2 con Eviews

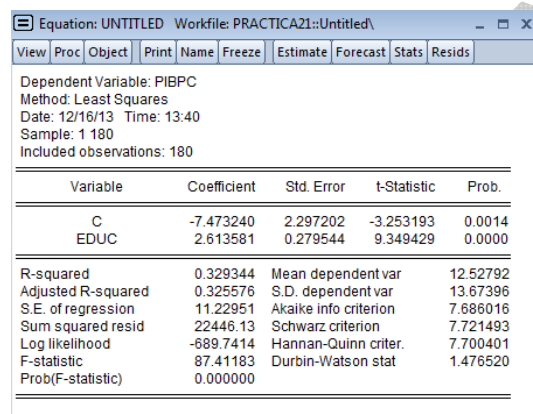
El siguiente modelo de regresión simple relaciona el PIB por habitante (variable *pibpc*) con la educación (variable *educ*) y otros factores no observados:

$$pibpc_i = \beta_0 + \beta_1 educ_i + \varepsilon_i$$

A partir del *Informe sobre el Desarrollo Humano* se han obtenido datos de *pibpc* (en miles de dólares PPA de 2005) y *educ* (en años de educación promedio) correspondientes a una muestra de 180 países para el año 2012 (fichero de datos: *practica2.wf1*). Responda a las siguientes cuestiones:

1) Obtenga la recta de regresión MCO e interprete los parámetros estimados.

Si usamos comandos, tenemos que escribir *ls pibpc c educ*



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.473240	2.297202	-3.253193	0.0014
EDUC	2.613581	0.279544	9.349429	0.0000

R-squared	0.329344	Mean dependent var	12.52792
Adjusted R-squared	0.325576	S.D. dependent var	13.67396
S.E. of regression	11.22951	Akaike info criterion	7.686016
Sum squared resid	22446.13	Schwarz criterion	7.721493
Log likelihood	-689.7414	Hannan-Quinn criter.	7.700401
F-statistic	87.41183	Durbin-Watson stat	1.476520
Prob(F-statistic)	0.000000		

$$pibpc = -7.47 + 2.61educ$$

Interpretación de $\hat{\beta}_0$: el valor estimado del PIB por habitante cuando los años de educación promedio valen cero es de -7.47 miles de dólares. Como vemos en este ejemplo, $\hat{\beta}_0$ no siempre tiene una interpretación económica razonable.

Interpretación de $\hat{\beta}_1$: el aumento estimado en el PIB por habitante cuando los años de formación aumentan en un año es de 2.61 miles dólares.

2) Calcule e interprete la media y la desviación típica del PIB por habitante

El PIB por habitante medio es de 12.52 miles de dólares y se puede obtener del cuadro de regresión en el apartado *Mean dependent var*.

La desviación típica se puede obtener del cuadro de regresión que aparece en el apartado 1. El apartado *S.D. dependent var* es la cuasi desviación típica del PIB por habitante y vale 13.67396. Por tanto, si llamamos y al PIB por habitante, tenemos

$$13.67396 = \sqrt{\frac{\sum_{i=1}^{180} (y_i - \bar{y})^2}{179}}.$$

Para obtener la desviación típica, basta con multiplicar esta expresión por $\sqrt{179}$ y dividirla por $\sqrt{180}$:

$$\sqrt{\frac{\sum_{i=1}^{180} (y_i - \bar{y})^2}{180}} = 13.67396 \frac{\sqrt{179}}{\sqrt{180}} = 13.636 \text{ miles de dólares}$$

La desviación típica es muy similar al valor medio del PIB por habitante, lo que indica un elevado grado de dispersión entre los niveles de PIB por habitante de los países de la muestra.

3) ¿En cuánto se estima que aumente el PIB por habitante promedio si los años de educación se incrementan en 1 año? En 2.61 miles de dólares.

¿Y si se incrementan en 4 años? En $2.61 \cdot 4 = 10.45$ miles de dólares.

4) ¿Cuál es el PIB por habitante estimado cuando $educ = 8$? Es de 13.43 miles de dólares.

$$pibpc = -7.47 + 2.61 \cdot 8 = 13.43$$

5) Calcule e interprete la suma total de cuadrados (STC), la suma de cuadrados de los residuos (SCE) y la suma de cuadrados de la regresión (SCR). Calcule e interprete el coeficiente de determinación.

En el cuadro de regresión que aparece en el apartado 1, *S.D. dependent var* es la cuasi desviación típica del PIB por habitante y vale 13.67396. Por tanto, si llamamos y al PIB por habitante, tenemos

$$13.67396 = \sqrt{\frac{\sum_{i=1}^{180} (y_i - \bar{y})^2}{179}}.$$

El numerador de esa expresión es la raíz cuadrada de STC, despejando se puede fácilmente ver que

$$STC = (13.67396)^2 \cdot 179 = 33468.91 (\text{miles de dólares})^2.$$

La suma de cuadrados de los residuos (SCE) aparece directamente en el cuadro de regresión en el apartado *Sum squared resid*. Por tanto, SCE=22446.13.

Para obtener la suma de cuadrados de la regresión (SCR), es importante tener en cuenta que como el modelo tiene constante se cumple $STC=SCE+SCR$. Por tanto,

$$SCR=STC-SCE=33468.91-22446.13=11022.78$$

El coeficiente de determinación se puede calcular a partir de las expresiones anteriores

$$R^2 = 1 - \frac{SCE}{STC} = 1 - \frac{22446.13}{33468.91} = 0.329,$$

aunque también podríamos haberlo obtenido directamente del cuadro de regresión en el apartado **R-squared**. Es la fracción de la variación muestral del PIB por habitante explicada por la función de regresión muestral, en este caso casi un 33%.

6) Considere ahora los siguientes modelos de regresión múltiple que relacionan el PIB por habitante con la educación y otros factores:

$$pibpc_i = \beta_0 + \beta_1 educ_i + \beta_2 tc_i + \varepsilon_i$$

$$pibpc_i = \beta_0 + \beta_1 educ_i + \beta_2 tc_i + \beta_3 mortal_i + \varepsilon_i$$

donde *tc* es la tasa de crecimiento anual del país en 2012 y *mortal* es la tasa de mortalidad de niños menores de cinco años (por cada 1.000 nacidos vivos).

a. Estime ambos modelos por MCO y compare las estimaciones y la bondad del ajuste de los tres modelos.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-18.48379	2.451867	-7.538660	0.0000
EDUC	3.368023	0.261534	12.87794	0.0000
TC	3.323891	0.431078	7.710657	0.0000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.01502	4.031231	-2.484357	0.0139
EDUC	2.615856	0.385353	6.788211	0.0000
TC	3.303628	0.424167	7.788503	0.0000
MORTAL	-0.062820	0.023959	-2.621987	0.0095

El coeficiente de *educ* cambia algo al introducir *tc* pero recupera su valor inicial cuando además se añade *mortal*. El coeficiente de *tc* no cambia mucho cuando se incluye *mortal*.

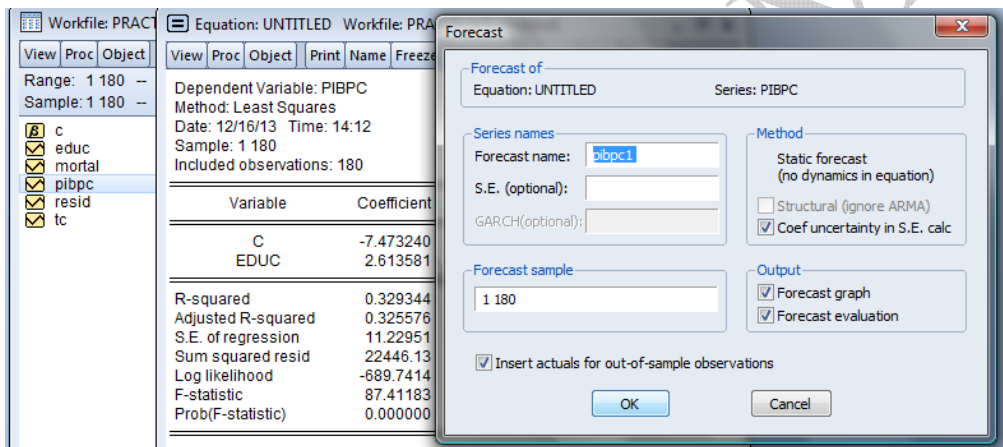
Al tratarse de modelos con diferente número de regresores (en concreto, cada uno de ellos contiene un regresor añadido a los regresores del modelo previo), para compararlos entre sí en términos de la bondad del ajuste, no es adecuado el uso del coeficiente de determinación R^2 . En su lugar debe emplearse el coeficiente de determinación ajustado, R_a^2 , cuyo valor aparece calculado en cada uno de los cuadros de

regresión, con la denominación *Adjusted R-squared*. El coeficiente de determinación ajustado pasa de 0.325 en la primera especificación a 0.492 en la segunda y a 0.509 en la tercera. En términos únicamente de la bondad del ajuste, esta última especificación sería por tanto la preferida.

b. Calcule el vector de PIB por habitante estimado para cada uno de los tres modelos y compárelos ¿Tienen las mismas medias aritméticas? ¿Por qué? Compárela con la media de la variable dependiente.

Vamos a describir cómo se calcularía el PIB por habitante estimado sólo para el primer modelo. En los otros dos modelos se han de seguir los mismos pasos.

Primero, estimamos la ecuación por MCO: *ls pibpc c educ* y una vez que tenemos el cuadro de regresión pinchamos en *Forecast*. Entonces nos pedirá que le demos un nombre a la variable dentro de la cual se meten las estimaciones del PIB por habitante según este primer modelo, que llamaremos *pibpc1*.



Una vez realizado el mismo procedimiento para los otros dos modelos, debemos tener las tres estimaciones, que llamaremos *pibpc1*, *pibpc2* y *pibpc3*. Para obtener información sobre estas variables, las seleccionamos y una vez seleccionadas, pinchamos con el botón derecho de ratón sobre ellas y las abrimos un grupo: *Open/as Group*. Una vez en el grupo, pinchamos en: *View/Descriptive Stats/Common Sample*

Group: UNTITLED Workfile: PRACTICA21::Untitled\									
View	Proc	Object	Print	Name	Freeze	Sample	Sheet	Stats	Spec
				PIBPC		PIBPC1		PIBPC2	PIBPC3
Mean				12.52792		12.52792		12.52792	12.52792
Median				7.638000		13.30473		13.06309	12.81200
Maximum				77.98700		27.28739		56.62592	58.79331
Minimum				0.329000		-4.336942		-6.464824	-8.039336
Std. Dev.				13.67396		7.847284		9.649346	9.830495
Skewness				1.766932		-0.291838		1.014477	0.898848
Kurtosis				6.703234		2.092101		6.595010	7.082155
Jarque-Bera				196.5160		8.737190		127.8057	149.2177
Probability				0.000000		0.012669		0.000000	0.000000
Sum				2255.025		2255.025		2255.025	2255.025
Sum Sq. Dev.				33468.93		11022.80		16666.67	17298.32
Observations				180		180		180	180

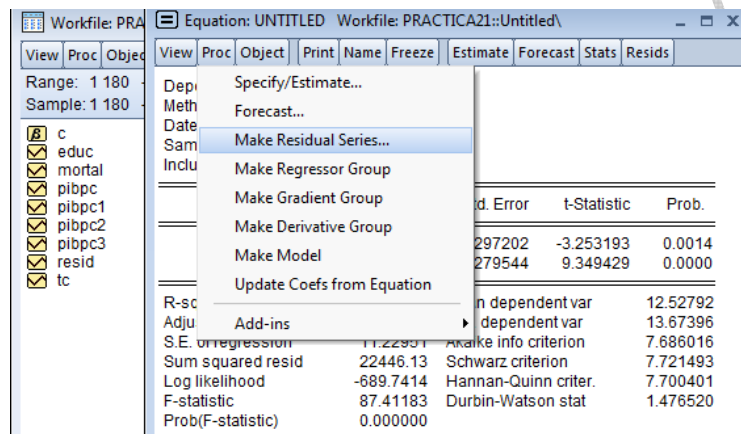
¿Tienen las mismas medias aritméticas? Sí, valen 12.52 miles de dólares.

¿Por qué? **Compárela con la media de la variable dependiente.** Porque estamos estimando modelos por MCO con constante y hemos demostrado que en estas estimaciones las medias muestrales de las variables estimadas coinciden con la media muestral de la variable dependiente.

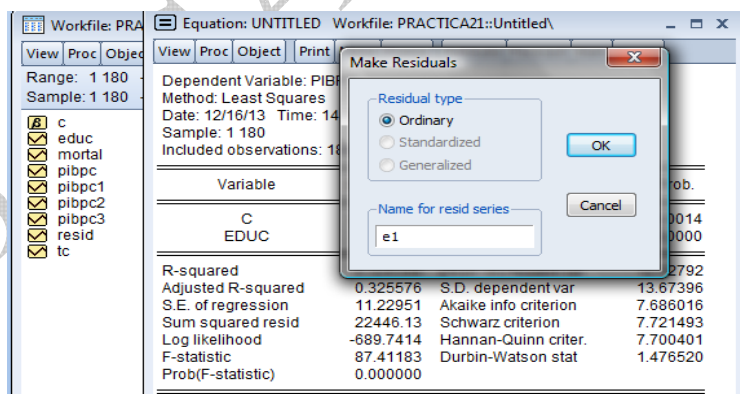
c. Calcule el vector de residuos de cada uno de los tres modelos, así como la media muestral de cada una de las series de residuos ¿Es la misma? ¿Por qué?

Vamos a describir cómo se calcularían los residuos sólo para el primer modelo. En los otros dos modelos se han de seguir los mismos pasos.

Primero, estimamos la ecuación por MCO: $ls\ pibpc\ c\ educ$ y una vez que tenemos el cuadro de regresión pinchamos en **Proc/Make Residual Series...**



Entonces nos pedirá que le demos un nombre a la variable dentro de la cual se meten los residuos de este primer modelo, que llamaremos $e1$.



Una vez realizado el mismo procedimiento para los otros dos modelos, debemos tener los tres residuos, que habremos llamado $e1$, $e2$ y $e3$. Para obtener información sobre estas variables, las seleccionamos y una vez seleccionadas, pinchamos con el botón derecho de ratón sobre ellas y las abrimos un grupo: **Open/as Group**.

Una vez en el grupo, pinchamos en: **View/Descriptive Stats/Common Sample**

G Group: UNTITLED Workfile: PRACTICA21::Untitled\			
View	Proc	Object	Print Name
E1 E2 E3			
Mean	-3.80E-15	-2.11E-15	-2.09E-15
Median	-1.972029	-1.096649	-1.778175
Maximum	66.38110	45.94559	45.30472
Minimum	-19.32509	-28.72581	-29.27109
Std. Dev.	11.19810	9.688518	9.504663
Skewness	2.256628	1.051046	1.075250
Kurtosis	11.76530	6.175788	6.093019
Jarque-Bera	728.9995	108.7831	106.4356
Probability	0.000000	0.000000	0.000000
Sum	-6.79E-13	-2.98E-13	-3.45E-13
Sum Sq. Dev.	22446.13	16802.26	16170.61
Observations	180	180	180

La media muestral debe ser matemáticamente cero en los tres casos porque hemos demostrado que cuando estimamos un modelo con constante por MCO, sus residuos suman cero y por tanto su media es cero. En la práctica, como el Eviews hace muchos cálculos sin incluir todos los decimales, vemos que la media estimada es casi cero en los tres casos.

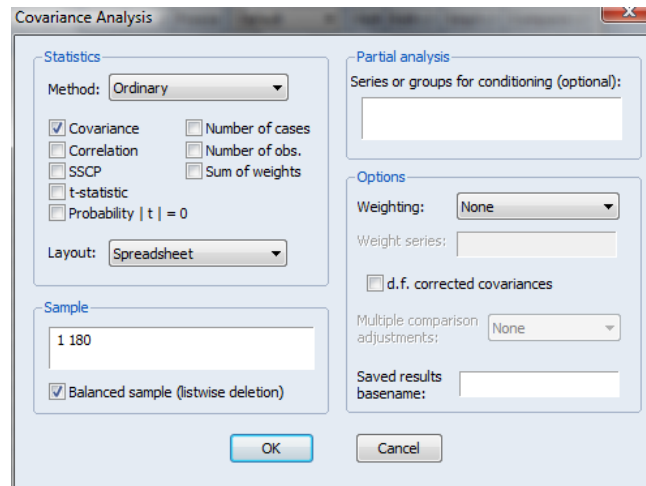
d. Calcule las covarianzas de los residuos con la variable dependiente PIB por habitante ¿Le parece lógico el resultado? ¿Por qué?

Podemos obtener la matriz de varianzas-covarianzas mediante ventanas seleccionado conjuntamente *e1 e2 e3* y *plibpc*, pinchando sobre ellos y con el botón derecho **Open/as Group**, y desde la ventana en la que aparecen ambas series seleccionamos **View/Covariance Analysis**:

Workfile: PRACTICA21 - (c:\...)			
View	Proc	Object	Save Freez
Range: 1 180 -- 180 obs Sample: 1 180 -- 180 obs			
<input checked="" type="checkbox"/>		c	
<input checked="" type="checkbox"/>		e1	
<input checked="" type="checkbox"/>		e2	
<input checked="" type="checkbox"/>		e3	
<input checked="" type="checkbox"/>		educ	
<input checked="" type="checkbox"/>		mortal	
<input checked="" type="checkbox"/>		plibpc	
<input checked="" type="checkbox"/>		plibpc1	
<input checked="" type="checkbox"/>		plibpc2	
<input checked="" type="checkbox"/>		plibpc3	
<input checked="" type="checkbox"/>		resid	
<input checked="" type="checkbox"/>		tc	

G Group: UNTITLED Workfile: PRACTICA21::Untitled\			
View	Proc	Object	Print Name
E2 E3 PIBPC			
Group Members	242	20.59170	46.98200
Spreadsheet	899	7.870682	34.54800
Dated Data Table	358	15.23943	42.48600
Graph...	617	15.85192	37.25100
Descriptive Stats	330	13.12022	34.43700
Covariance Analysis...	128	10.93729	35.64000
N-Way Tabulation...	681	12.00306	35.04800
Tests of Equality...	781	17.22117	37.97900
Principal Components...	472	10.51955	30.66000
Correlogram (1) ...	216	10.62928	35.71600
Cross Correlation (2) ...	777	5.874377	27.54100
	851	11.59832	33.61800
	538	11.19273	32.39900
	633	-1.677909	26.72000

Ahora elegimos la opción *Covariance*



El resultado que obtenemos en Eviews es el siguiente:

View	Proc	Object	Print	Name	Freeze	Sample	Sheet	Stats	Spec
Covariance									
		E1	E2	E3	PIBPC				
E1		124.7007	93.34589	89.83674	124.7007				
E2		93.34589	93.34589	89.83674	93.34589				
E3		89.83674	89.83674	89.83674	89.83674				
PIBPC		124.7007	93.34589	89.83674	185.9385				

Como vimos en las clases de teoría, los residuos no tienen covarianza nula con la variable dependiente sino con la variable dependiente estimada. Según el método MCO, tratamos de descomponer el PIB por habitante en dos componentes, una parte que podemos modelizar a través de las variables explicativas y otra que no podremos modelizar que se queda como un ruido. Cuanto más seamos capaces de modelizar, menos se relacionará el ruido con la variable dependiente. En este caso, aunque el ruido no es observable, sí que podemos ver como la covarianza entre los residuos y el PIB por habitante disminuye conforme el modelo es más completo y la parte que podemos modelizar del PIB por habitante está mejor controlada a través de variables explicativas.