

Solución a la práctica 4.1 con Eviews

Se pretende explicar la demanda de turismo nacional en las 52 provincias españolas durante el año 2014. Para ello se propone el siguiente modelo econométrico:

$$\log(\text{turismo}_i) = \beta_0 + \beta_1 \log(\text{densidad}_i) + \beta_2 \log(\text{hoteles}_i) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \beta_5 \text{pibpc}_i + \beta_6 \text{paro}_i + \varepsilon_i$$

donde *turismo* es el número de pernoctaciones hoteleras de turistas nacionales en cada una de las provincias españolas en 2014, *densidad* es la densidad de población de cada provincia (población según censo a 1 de enero de 2015), *hoteles* es el número de establecimientos hoteleros en la provincia de destino en 2014, *kmco* son los kilómetros de costa de cada provincia, *ipc* es el IPC provincial en 2014 (base 2011), *pibpc* es el índice del PIB per cápita provincial en 2013 (España=100), y *paro* es la tasa de paro (%) de cada provincia en 2014 según la EPA.¹

Estime el modelo por MCO y responda a las siguientes preguntas (fichero de datos: *practica41.wf1*):

Para estimar el modelo del enunciado escribimos en la ventana de comandos:

ls log(turismo) c log(densidad) log(hoteles) kmco ipc pibpc paro

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	32.22926	12.19307	2.643244	0.0113
LOG(DENSIDAD)	0.536358	0.115566	4.641129	0.0000
LOG(HOTELES)	0.525447	0.130980	4.011669	0.0002
KMCO	0.000546	0.000232	2.355016	0.0229
IPC	-0.213179	0.116741	-1.826077	0.0745
PIBPC	0.005052	0.004918	1.027235	0.3098
PARO	0.015981	0.017079	0.935739	0.3544

R-squared	0.858940	Mean dependent var	14.03099
Adjusted R-squared	0.840132	S.D. dependent var	1.040792
S.E. of regression	0.416145	Akaike info criterion	1.209084
Sum squared resid	7.792953	Schwarz criterion	1.471751
Log likelihood	-24.43618	Hannan-Quinn criter.	1.309784
F-statistic	45.66887	Durbin-Watson stat	1.920552
Prob(F-statistic)	0.000000		

¹ Fuente de Datos: Instituto Nacional de Estadística (INE) e Instituto Geográfico Nacional.

1) Realice una estimación de la varianza de ε mediante MCO.

El estimador de la varianza de ε mediante MCO es:

$$s^2 = \frac{\sum_{i=1}^N e_i^2}{N-K}$$

El numerador lo podemos obtener del cuadro de regresión en “Sum squared resid” y vale 7.7929 y el denominador es $N=52$ observaciones menos $K=7$ parámetros que estimamos. Por tanto la varianza estimada es

$$s^2 = \frac{7.7929}{52-7} = 0.1732$$

Este dato también se podría obtener del cuadro de regresión en *S.E. of regression* que es la estimación de la desviación típica estimada, es decir $\sqrt{s^2}$, y vale 0.41614. Por tanto, la varianza estimada será $0.41614^2=0.1732$.

2) ¿Cuál es la estimación de la varianza de los estimadores MCO de los parámetros del modelo?

Eviews calcula automáticamente las varianzas estimadas de los estimadores MCO. Para ello en el output de estimación del modelo seleccionamos *View/Covariance Matrix*

Coefficient Covariance Matrix							
	C	LOG(DENSI...	LOG(HOTEL...	KMCO	IPC	PIBPC	PARO
C	148.6710	0.113152	0.342648	-0.000490	-1.416856	-0.012895	-0.102920
LOG(DENSI...	0.113152	0.013356	-0.011453	-1.63E-06	-0.000121	-0.000202	-0.000910
LOG(HOTEL...	0.342648	-0.011453	0.017156	-7.23E-06	-0.004379	0.000101	0.000506
KMCO	-0.000490	-1.63E-06	-7.23E-06	5.38E-08	4.92E-06	7.51E-08	2.06E-07
IPC	-1.416856	-0.000121	-0.004379	4.92E-06	0.013629	8.24E-05	0.000842
PIBPC	-0.012895	-0.000202	0.000101	7.51E-08	8.24E-05	2.42E-05	6.40E-05
PARO	-0.102920	-0.000910	0.000506	2.06E-07	0.000842	6.40E-05	0.000292

Los elementos de la diagonal principal son las varianzas estimadas de los coeficientes estimados $\hat{V}(\hat{\beta}_j)$. Por ejemplo, los resultados de Eviews indican que $\hat{V}(\hat{\beta}_1) = 0.0134$ y $\hat{V}(\hat{\beta}_2) = 0.0172$. Los elementos fuera de la diagonal principal son las covarianzas estimadas.

En la columna *Std. Error* del output de estimación, Eviews muestra la desviación típica estimada de los coeficientes ($\sqrt{\hat{V}(\hat{\beta}_j)}$).

3) Contraste la significatividad conjunta de las variables incluidas en el modelo.

Las hipótesis del contraste son:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \text{no } H_0$$

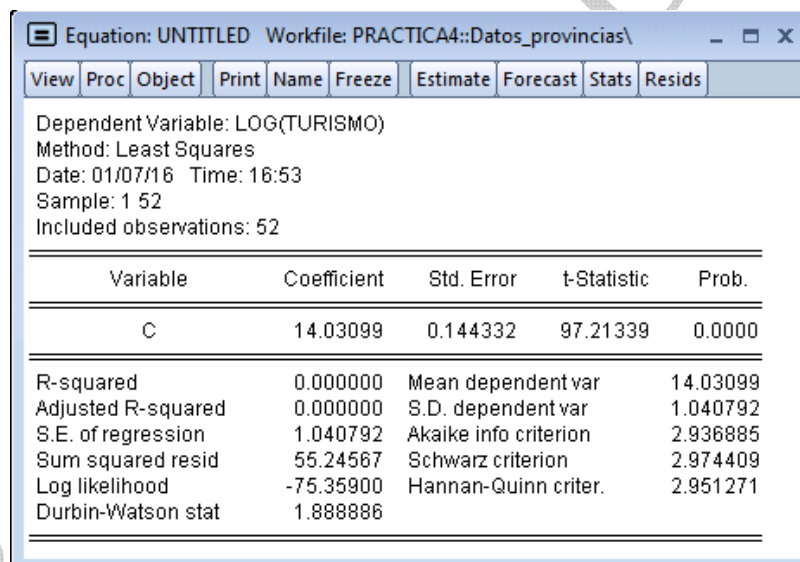
Es decir, bajo la alternativa, por lo menos uno de los coeficientes de los regresores del modelo difiere de 0.

El estadístico de contraste es:

$$F = \frac{(SCE_R - SCE_{NR})/q}{SCE_{NR}/(N - K)} \sim F_{q, N-K}$$

En este caso, el **modelo restringido** es: $\log(\text{turismo}_i) = \beta_0 + \varepsilon_i$, y la estimación es:

ls log(turismo) c



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.03099	0.144332	97.21339	0.0000

R-squared	0.000000	Mean dependent var	14.03099
Adjusted R-squared	0.000000	S.D. dependent var	1.040792
S.E. of regression	1.040792	Akaike info criterion	2.936885
Sum squared resid	55.24567	Schwarz criterion	2.974409
Log likelihood	-75.35900	Hannan-Quinn criter.	2.951271
Durbin-Watson stat	1.888886		

Entonces, el estadístico de contraste se calcula como:

$$F = \frac{(55.2457 - 7.7929) / 6}{7.7929 / (52 - 7)} = 45.669$$

La regla de rechazo es: $F > F_{q, N-K; \alpha} \Rightarrow RH_0$

Utilizando las tablas de la distribución F de Snedecor obtenemos:

$$\text{Valor crítico al 5\%: } F_{6,45;0.05} \approx F_{6,50;0.05} = 2.286$$

$$\text{Valor crítico al 10\%: } F_{6,45;0.1} \approx F_{6,50;0.1} = 1.895$$

A los niveles de significación habituales del 5% y 10% los valores críticos para este contraste son aproximadamente 2.286 y 1.895, respectivamente. Estos valores se

pueden calcular de forma exacta mediante las funciones de Eviews @*qfdist*(0.95,6,45) y @*qfdist*(0.90,6,45), respectivamente.

Como el estadístico F calculado (45.669) es mayor que los valores críticos al 5% y al 10%, se puede rechazar la hipótesis nula. Concluimos que al menos una de las variables explicativas es relevante, y que el modelo es estadísticamente significativo.

Si observamos el output de estimación del modelo no restringido en la página 1, veremos que el valor de este estadístico F aparece en la parte inferior izquierda bajo la denominación de F-Statistic. Además Eviews presenta también el p-valor asociado, 0.000000 en este caso, indicando que se RH_0 para cualquier nivel de significatividad.

4) Contraste la significatividad individual de las variables incluidas en el modelo ¿Está de acuerdo con el modelo propuesto? Si no es así, ¿qué modelo propone?

Para contrastar la relevancia estadística de cada una de las variables incluidas en el modelo, debemos realizar un contraste de significatividad individual de su coeficiente. Las hipótesis del contraste son:

$$H_0 : \beta_j = 0$$

$$H_A : \beta_j \neq 0$$

donde $j=1, 2, 3, 4, 5, 6$

El estadístico de contraste es:
$$t = \frac{\hat{\beta}_j}{\sqrt{\hat{V}\text{ar}(\hat{\beta}_j)}} \underset{H_0}{\sim} t_{N-K}$$

La regla de rechazo es: $|t| > t_{52-7; \alpha/2} \Rightarrow RH_0$

Utilizando las tablas de la distribución t de Student:

$$\text{Valor crítico al 5\%: } t_{45; 0.025} \approx t_{50; 0.025} = 2.009$$

$$\text{Valor crítico al 10\%: } t_{45; 0.05} \approx t_{50; 0.05} = 1.676$$

A los niveles de significación habituales del 5% y 10% los valores críticos para este contraste son aproximadamente 2.009 y 1.676, respectivamente. Estos valores se pueden calcular de forma exacta mediante las funciones de Eviews @*qtdist*(0.975,45) y @*qtdist*(0.95,45), respectivamente.

Con los datos de la estimación se obtiene:

- El estadístico t para β_1 es $t = \frac{0.5364}{0.1156} = 4.640$, por lo que su estimación es estadísticamente significativa, de tal manera que la densidad de población afecta a la demanda turística.
- El estadístico t para β_2 es $t = \frac{0.5254}{0.1309} = 4.014$, por lo que su estimación es estadísticamente significativa, de tal manera que el número de hoteles afecta a la demanda turística.

- El estadístico t para β_3 es $t = \frac{0.000546}{0.000232} = 2.353$, por lo que su estimación es estadísticamente significativa, de tal manera que los kilómetros de costa que tiene cada provincia afectan a la demanda turística.
- El estadístico t para β_4 es $t = \frac{-0.2132}{0.1167} = -1.827$, por lo que su estimación es estadísticamente significativa al 10% pero no al 5%.
- El estadístico t para β_5 es $t = \frac{0.00505}{0.00492} = 1.026$, por lo que su estimación no es estadísticamente significativo ni al 10% ni al 5%, y el PIB per cápita provincial no sería una variable relevante para explicar el número de pernoctaciones hoteleras en la provincia.
- El estadístico t para β_6 es $t = \frac{0.01598}{0.01708} = 0.936$, por lo que su estimación no es estadísticamente significativa ni al 10% ni al 5%, y la tasa de paro de la provincia no sería relevante para explicar el número de pernoctaciones hoteleras en la provincia.

Si observamos el output de estimación del modelo en la página 1, vemos que el valor de este estadístico t aparece en la columna *t-Statistic*. Además Eviews presenta también el p-valor asociado a este estadístico t en la columna *Prob*:

- $\log(\text{densidad})$ tiene un p-valor asociado igual a 0.000, por tanto es relevante para cualquier nivel de significatividad.
- $\log(\text{hoteles})$ tiene un p-valor asociado igual a 0.000, por tanto es relevante para cualquier nivel de significatividad.
- $kmco$ tiene un p-valor asociado igual a 0.0229, por tanto es relevante al 5% pero no al 1%.
- ipc tiene un p-valor asociado igual a 0.0745, por tanto es relevante al 10% pero no al 5%.
- $pibpc$ tiene un p-valor asociado igual a 0.3098, por tanto no es relevante ni al 5% ni al 10%.
- $paro$ tiene un p-valor asociado igual a 0.3544, por tanto no es relevante ni al 5% ni al 10%.

El modelo propuesto no es correcto porque presenta inclusión de variables irrelevantes, lo que supone una pérdida de precisión en las estimaciones. Para obtener la especificación adecuada hay que eliminar una a una las variables irrelevantes, comenzando por la que tiene un mayor p-valor. En nuestro caso eliminamos en primer lugar la variable *paro*, y volvemos a estimar el modelo. El resultado que se obtiene es:

ls (turismo) c log(densidad) log(hoteles) kmco ipc pibpc

Equation: UNTITLED Workfile: PRACTICA4::Datos_provincias\									
View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: LOG(TURISMO)									
Method: Least Squares									
Date: 01/07/16 Time: 16:57									
Sample: 1 52									
Included observations: 52									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	37.86820	10.58547	3.577374	0.0008					
LOG(DENSIDAD)	0.586207	0.102417	5.723740	0.0000					
LOG(HOTELES)	0.497705	0.127408	3.906399	0.0003					
KMCO	0.000535	0.000231	2.312578	0.0253					
IPC	-0.259294	0.105686	-2.453440	0.0180					
PIBPC	0.001546	0.003181	0.485929	0.6293					
R-squared	0.856195	Mean dependent var	14.03099						
Adjusted R-squared	0.840564	S.D. dependent var	1.040792						
S.E. of regression	0.415582	Akaike info criterion	1.189893						
Sum squared resid	7.944588	Schwarz criterion	1.415037						
Log likelihood	-24.93723	Hannan-Quinn criter.	1.276208						
F-statistic	54.77565	Durbin-Watson stat	1.949958						
Prob(F-statistic)	0.000000								

La variable *piGPC* sigue sin ser relevante (p -valor=0.6293), por tanto hay que eliminarla del modelo. El modelo resultante es:

$$\log(\text{turismo}_i) = \beta_0 + \beta_1 \log(\text{densidad}_i) + \beta_2 \log(\text{hoteles}_i) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \varepsilon_i$$

La estimación MCO de este modelo es:

ls (turismo) c log(densidad) log(hoteles) kmco ipc

Equation: UNTITLED Workfile: PRACTICA4::Datos_provincias\									
View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: LOG(TURISMO)									
Method: Least Squares									
Date: 01/07/16 Time: 17:00									
Sample: 1 52									
Included observations: 52									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	36.39294	10.05802	3.618299	0.0007					
LOG(DENSIDAD)	0.586636	0.101577	5.775267	0.0000					
LOG(HOTELES)	0.499253	0.126328	3.952025	0.0003					
KMCO	0.000531	0.000229	2.313660	0.0251					
IPC	-0.243720	0.099888	-2.439947	0.0185					
R-squared	0.855457	Mean dependent var	14.03099						
Adjusted R-squared	0.843156	S.D. dependent var	1.040792						
S.E. of regression	0.412191	Akaike info criterion	1.156552						
Sum squared resid	7.985369	Schwarz criterion	1.344172						
Log likelihood	-25.07035	Hannan-Quinn criter.	1.228481						
F-statistic	69.54074	Durbin-Watson stat	1.936110						
Prob(F-statistic)	0.000000								

En este modelo todas las variables son relevantes al 5%. Este es el modelo que proponemos para estudiar la demanda de noches de hotel.

5) ¿Cuál es la interpretación de β_2 y de su estimación?

El coeficiente β_2 es la elasticidad de la demanda de noches de hotel respecto al número de establecimientos hoteleros y mide aproximadamente el cambio en la variable *turismo* en % por cada variación de un 1% en *hoteles*, manteniendo fijas el resto de variables del modelo.

Según los resultados, en promedio, se estima que un incremento de un 1% en el número de hoteles de una provincia produce aproximadamente un incremento de un 0.499 % en las pernoctaciones hoteleras en dicha provincia.

6) Contraste la hipótesis nula de que un incremento de un 1% en el número de establecimientos hoteleros produce, en promedio, un incremento de la misma cuantía en las pernoctaciones frente a la alternativa de que dicho incremento sea distinto.

Las hipótesis del contraste son:

$$H_0 : \beta_2 = 1$$

$$H_1 : \beta_2 \neq 1$$

El estadístico de contraste es: $t = \frac{\hat{\beta}_2 - 1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}} \underset{H_0}{\sim} t_{N-K}$

La regla de rechazo es: $|t| > t_{47; \alpha/2} \Rightarrow RH_0$

A los niveles de significación habituales del 5% y 10% los valores críticos para este contraste son aproximadamente 2.009 y 1.676, respectivamente.

Con los datos de la estimación el resultado es:

$$t = \frac{0.4993 - 1}{0.1263} = -3.964$$

El estadístico t es -3.964 por lo que se RH_0 .

7) Contraste la hipótesis nula de que un incremento de un 1% en el número de establecimientos hoteleros produce, en promedio, un incremento de la misma cuantía en las pernoctaciones frente a la alternativa de que dicho incremento sea menor.

Las hipótesis del contraste son:

$$H_0 : \beta_2 = 1$$

$$H_1 : \beta_2 < 1$$

El estadístico de contraste es el mismo que en el apartado anterior:

$$t = \frac{0.4993 - 1}{0.1263} = -3.964$$

Este es un contraste unilateral por lo que la región crítica está en una sola cola de la distribución, en este caso en la cola de la izquierda. La regla de rechazo es:

$$t < -t_{47;\alpha} \Rightarrow RH_0$$

$$\text{Valor crítico al 5\%: } -t_{47;0.05} \approx -t_{50;0.05} = -1.676$$

$$\text{Valor crítico al 10\%: } -t_{47;0.1} \approx -t_{50;0.1} = -1.299$$

Dado que el estadístico t es -3.964 se RH_0 . Un incremento de un 1% en *hoteles* produce, en promedio, un incremento de menor cuantía en la demanda hotelera.

8) Construya un intervalo de confianza para β_2 al 95% e interprételo.

El intervalo de confianza es:

$$\left(\hat{\beta}_2 \pm t_{N-K;\alpha/2} \sqrt{\hat{V}ar(\hat{\beta}_2)} \right) = (0.4993 \pm 2.009 \times 0.1263) = (0.246, 0.753)$$

donde $\alpha = 100 - 95\%$.

Es el intervalo de posibles valores para el valor teórico β_2 , con un nivel de confianza del 95%.

El intervalo de confianza también se puede calcular **automáticamente con Eviews**. Eviews calcula el intervalo de confianza de todos los parámetros del modelo para tres niveles de confianza (90%, 95% y 99%). En el cuadro de estimación seleccionamos *View/Coefficient Diagnostics/Confidence Intervals*

Coefficient Confidence Intervals

Date: 05/17/16 Time: 14:27

Sample: 1 52

Included observations: 52

Variable	Coefficient	90% CI		95% CI		99% CI	
		Low	High	Low	High	Low	High
C	36.39294	19.51631	53.26956	16.15880	56.62707	9.391616	63.39426
LOG(DENSIDAD)	0.586636	0.416197	0.757076	0.382289	0.790984	0.313946	0.859327
LOG(HOTELES)	0.499253	0.287283	0.711223	0.245113	0.753394	0.160118	0.838389
KMCO	0.000531	0.000146	0.000915	6.92E-05	0.000992	-8.50E-05	0.001146
IPC	-0.243720	-0.411324	-0.076116	-0.444668	-0.042773	-0.511874	0.024433

9) Contraste conjuntamente las hipótesis de que la elasticidad de la demanda respecto al número de hoteles es idéntica a la elasticidad de la demanda respecto a la densidad de población, y que por cada punto que se incrementa el

IPC provincial se espera una disminución del 50% de las pernoctaciones. Realice el contraste de forma manual y automática.

Como las variables *turismo*, *densidad* y *hoteles* están en logaritmos, β_1 y β_2 son elasticidades. El IPC no está en logaritmos, por lo que β_4 es una "semielasticidad", y para interpretar el efecto de *ipc* sobre *turismo* en términos porcentuales hay que multiplicar β_4 por 100. Por tanto, las hipótesis del contraste en términos de los parámetros del modelo son:

$$H_0 : \beta_1 = \beta_2$$

$$\beta_4 = -0.5$$

$$H_1 : \text{no } H_0$$

Bajo la alternativa, por lo menos una de las dos restricciones no se cumple. El estadístico de contraste es

$$F = \frac{(\text{SCE}_R - \text{SCE}_{NR})/q}{\text{SCE}_{NR}/(N-K)} \sim F_{q, N-K}$$

El modelo restringido es:

$$\log(\text{turismo}_i) = \beta_0 + \beta_2 \log(\text{densidad}_i) + \beta_2 \log(\text{hoteles}_i) + \beta_3 \text{kmco}_i - 0.5 \cdot \text{ipc}_i + \varepsilon_i \Rightarrow$$

$$\log(\text{turismo}_i) + 0.5 \cdot \text{ipc}_i = \beta_0 + \beta_2 (\log(\text{densidad}_i) + \log(\text{hoteles}_i)) + \beta_3 \text{kmco}_i + \varepsilon_i$$

Estimamos el modelo restringido:

*ls log(turismo)+0.5*ipc c log(densidad)+log(hoteles) kmco*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	62.50179	0.214197	291.7956	0.0000
LOG(DENSIDAD)+LOG(HOTELES)	0.582086	0.039038	14.91066	0.0000
KMCO	0.000423	0.000235	1.795027	0.0788

R-squared	0.856769	Mean dependent var	65.80345
Adjusted R-squared	0.850923	S.D. dependent var	1.120538
S.E. of regression	0.432645	Akaike info criterion	1.218162
Sum squared resid	9.171892	Schwarz criterion	1.330733
Log likelihood	-28.67220	Hannan-Quinn criter.	1.261319
F-statistic	146.5529	Durbin-Watson stat	2.111939
Prob(F-statistic)	0.000000		

El valor del estadístico F es:

$$F = \frac{(9.1719 - 7.9854) / 2}{7.9854 / (52 - 5)} = 3.4918$$

La regla de rechazo es: $F > F_{2,47;\alpha} \Rightarrow RH_0$

Utilizando las tablas de la distribución F de Snedecor:

Valor crítico al 5%: $F_{2,47;0.05} \approx F_{2,50;0.05} = 3.183$

Valor crítico al 10%: $F_{2,47;0.1} \approx F_{2,50;0.1} = 2.412$

Los valores críticos al 5% y al 10% son aproximadamente 3.183 y 2.412, respectivamente. Por tanto, se rechaza la hipótesis nula.

De forma automática con Eviews, en el output de estimación del modelo no restringido, seleccionamos *View/Coefficient Diagnostics/Wald Test-Coefficient Restrictions...* En el cuadro que se abre escribimos $c(2)=c(3)$, $c(5)=-0.5$. El resultado se presenta en la siguiente tabla:

Test Statistic	Value	df	Probability
F-statistic	3.491799	(2, 47)	0.0386
Chi-square	6.983598	2	0.0304

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(3)	0.087383	0.213671
0.5 + C(5)	0.256280	0.099888

Restrictions are linear in coefficients.

Como vemos, el valor del estadístico F es 3.4918 y su p-valor 0.0386. Por tanto, se puede rechazar la hipótesis nula al 5%. Al menos una de las dos restricciones no se verifica.

10) Considere ahora la hipótesis nula de que la elasticidad de la demanda de noches de hotel respecto al número de hoteles es idéntica a la elasticidad de la demanda respecto a la densidad de población. Estime un modelo que proporcione directamente el estadístico t para contrastar dicha hipótesis nula ¿A qué conclusión se llega?

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 \\ H_0 : \beta_1 \neq \beta_2 \end{array} \quad \rightarrow \quad \begin{array}{l} H_0 : \beta_1 - \beta_2 = 0 \\ H_1 : \beta_1 - \beta_2 \neq 0 \end{array}$$

Reparametrizamos el modelo haciendo $\delta = \beta_1 - \beta_2$, con lo que $\beta_1 = \delta + \beta_2$. La hipótesis nula a contrastar es:

$$H_0 : \delta = 0$$

$$H_1 : \delta \neq 0$$

El modelo reparametrizado es:

$$\begin{aligned} \log(\text{turismo}_i) &= \beta_0 + (\delta + \beta_2) \log(\text{densidad}_i) + \beta_2 \log(\text{hoteles}_i) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \varepsilon_i \Rightarrow \\ \log(\text{turismo}_i) &= \beta_0 + \delta \log(\text{densidad}_i) + \beta_2 (\log(\text{densidad}_i) + \log(\text{hoteles}_i)) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \varepsilon_i \end{aligned}$$

Estimamos el modelo reparametrizado:

ls log(turismo) c log(densidad) log(densidad)+log(hoteles) kmco ipc

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.39294	10.05802	3.618299	0.0007
LOG(DENSIDAD)	0.087383	0.213671	0.408960	0.6844
LOG(DENSIDAD)+LOG(HOTELES)	0.499253	0.126328	3.952025	0.0003
KMCO	0.000531	0.000229	2.313660	0.0251
IPC	-0.243720	0.099888	-2.439947	0.0185

R-squared	0.855457	Mean dependent var	14.03099
Adjusted R-squared	0.843156	S.D. dependent var	1.040792
S.E. of regression	0.412191	Akaike info criterion	1.156552
Sum squared resid	7.985369	Schwarz criterion	1.344172
Log likelihood	-25.07035	Hannan-Quinn criter.	1.228481
F-statistic	69.54074	Durbin-Watson stat	1.936110
Prob(F-statistic)	0.000000		

El valor del estadístico t para el contraste de significatividad individual de δ es 0.4089 y el p-valor 0.6844, por lo que ahora no se puede rechazar la hipótesis nula a los niveles habituales del 5% o 10%.

11) Contraste la hipótesis nula del apartado 10) utilizando el estadístico F. Realice el contraste de forma manual y automática.

$$H_0 : \beta_1 = \beta_2$$

$$H_1 : \beta_1 \neq \beta_2$$

El estadístico de contraste es

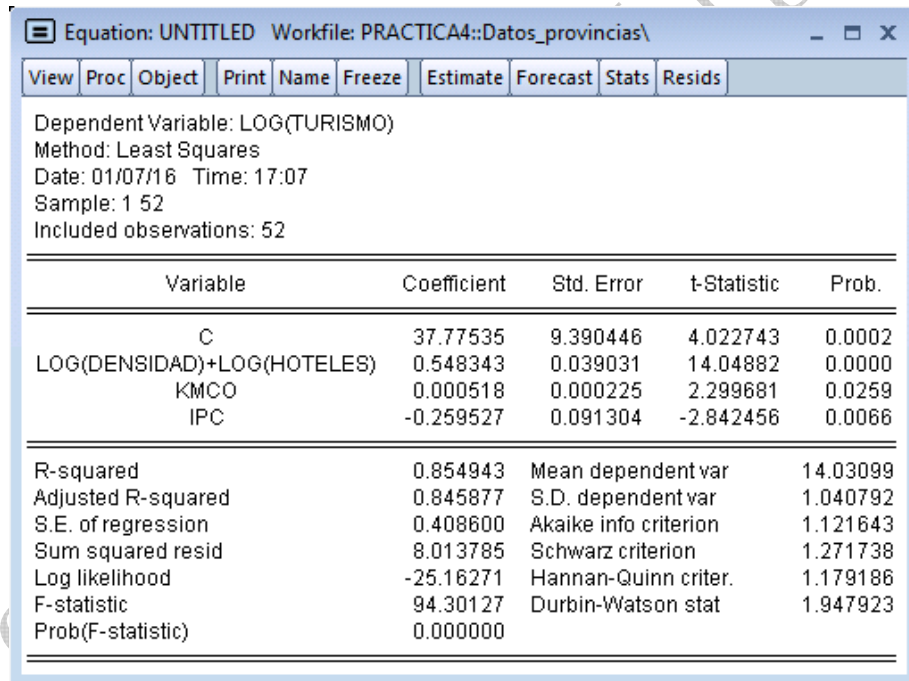
$$F = \frac{(\text{SCE}_R - \text{SCE}_{NR})/q}{\text{SCE}_{NR}/(N-K)} \sim F_{q, N-K}$$

El modelo restringido es:

$$\log(\text{turismo}_i) = \beta_0 + \beta_2 (\log(\text{densidad}_i) + \log(\text{hoteles}_i)) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \varepsilon_i$$

Estimamos el modelo restringido:

ls log(turismo) c log(densidad)+log(hoteles) kmco ipc



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	37.77535	9.390446	4.022743	0.0002
LOG(DENSIDAD)+LOG(HOTELES)	0.548343	0.039031	14.04882	0.0000
KMCO	0.000518	0.000225	2.299681	0.0259
IPC	-0.259527	0.091304	-2.842456	0.0066

R-squared	0.854943	Mean dependent var	14.03099
Adjusted R-squared	0.845877	S.D. dependent var	1.040792
S.E. of regression	0.408600	Akaike info criterion	1.121643
Sum squared resid	8.013785	Schwarz criterion	1.271738
Log likelihood	-25.16271	Hannan-Quinn criter.	1.179186
F-statistic	94.30127	Durbin-Watson stat	1.947923
Prob(F-statistic)	0.000000		

El valor del estadístico F es:

$$F = \frac{(8.0138 - 7.9854)/1}{7.9854/(52 - 5)} = 0.1672$$

La regla de rechazo es: $F > F_{1,52-5} \Rightarrow RH_0$

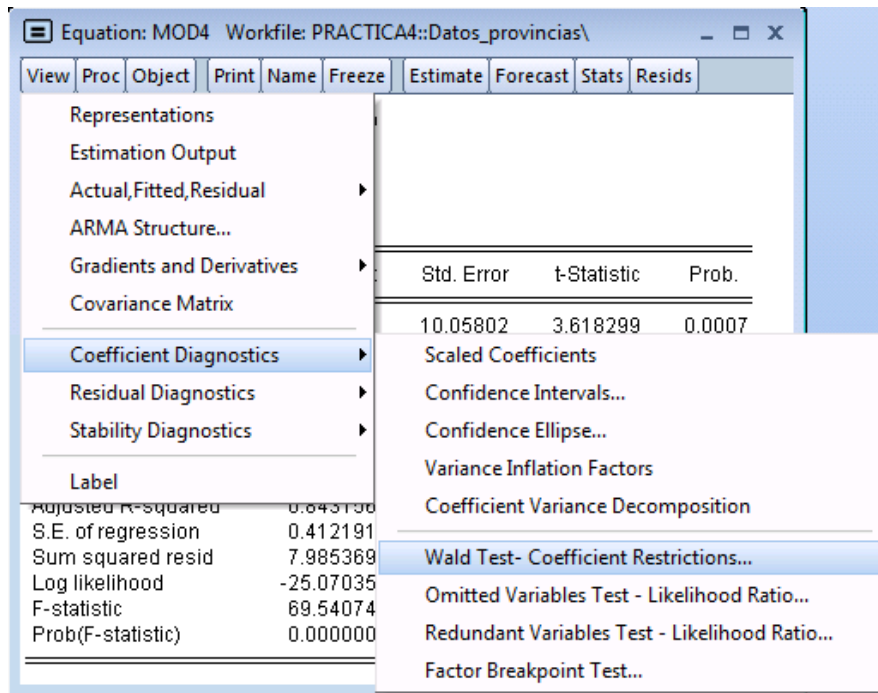
Utilizando las tablas de la distribución F de Snedecor:

$$\text{Valor crítico al 5\%: } F_{1,47;0.05} \approx F_{1,50;0.05} = 4.034$$

$$\text{Valor crítico al 10\%: } F_{1,47;0.1} \approx F_{1,50;0.1} = 2.809$$

Los valores críticos al 5% y al 10% son aproximadamente 4.034 y 2.809, respectivamente, por lo que no se puede rechazar la hipótesis nula.

Eviews puede realizar cualquier contraste bilateral de forma automática. Para ello en el output de estimación del modelo no restringido, seleccionamos *View/Coefficient Diagnostics/Wald Test-Coefficient Restrictions...*



En el cuadro que se abre escribimos $c(2)=c(3)$ (Nota: $c(1)$ se corresponde con el primer coeficiente estimado, en nuestro caso con β_0 ; $c(2)$ con el segundo, etc.). El resultado se presenta en la siguiente tabla:

The screenshot shows the 'Wald Test: Equation: Untitled' dialog box. The test results are as follows:

Test Statistic	Value	df	Probability
t-statistic	0.408960	47	0.6844
F-statistic	0.167249	(1, 47)	0.6844
Chi-square	0.167249	1	0.6826

Null Hypothesis: C(2)=C(3)
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(3)	0.087383	0.213671

Restrictions are linear in coefficients.

Como vemos, el valor del estadístico F es 0.1672 y su p-valor 0.6844. Por lo tanto, no se rechaza la hipótesis nula.

Como la hipótesis nula no se rechaza conviene incorporar la restricción en el modelo y continuar la práctica con el modelo en el que se impone la restricción $\beta_1 = \beta_2$. La estimación de este "modelo restringido" es más precisa que la del modelo sin restringir.

12) Lleve a cabo una predicción del logaritmo de las pernoctaciones hoteleras para una provincia hipotética que está situada en el interior, cuya densidad de población es igual a 4, cuenta con 300 establecimientos hoteleros, tiene una tasa de paro del 20%, un índice de PIB per cápita del 90% sobre la media nacional y presenta un IPC igual a 103. ¿Cuál es la predicción para el número de pernoctaciones hoteleras? Obtenga un intervalo de confianza para cada una de las predicciones al 95%.

La predicción se lleva a cabo utilizando el "modelo restringido" estimado en el apartado anterior:

$$\log(\text{turismo}_i) = \beta_0 + \beta_2 (\log(\text{densidad}_i) + \log(\text{hoteles}_i)) + \beta_3 \text{kmco}_i + \beta_4 \text{ipc}_i + \varepsilon_i$$

Predicción del logaritmo de las pernoctaciones hoteleras: $\log(\text{turismo})$

Para llevar a cabo la predicción y obtener la desviación típica necesaria para calcular el intervalo de confianza, estimamos el siguiente modelo reparametrizado:

$$\log(\text{turismo}_i) = \theta + \beta_2 (\log(\text{densidad}_i) + \log(\text{hoteles}_i) - \log(4) - \log(300)) + \beta_3 \text{kmco}_i + \beta_4 (\text{ipc}_i - 103) + \varepsilon_i$$

ls log(turismo) c log(densidad)+log(hoteles)-log(4)-log(300) kmco ipc-103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.93182	0.120869	123.5375	0.0000
LOG(DENSIDAD)+LOG(HOTELES)-LOG(4)-LOG(300)	0.548343	0.039031	14.04882	0.0000
KMCO	0.000518	0.000225	2.299681	0.0259
IPC-103	-0.259527	0.091304	-2.842456	0.0066

R-squared	0.854943	Mean dependent var	14.03099
Adjusted R-squared	0.845877	S.D. dependent var	1.040792
S.E. of regression	0.408600	Akaike info criterion	1.121643
Sum squared resid	8.013785	Schwarz criterion	1.271738
Log likelihood	-25.16271	Hannan-Quinn criter.	1.179186
F-statistic	94.30127	Durbin-Watson stat	1.947923
Prob(F-statistic)	0.000000		

Del output de Eviews se puede obtener el valor predicho para el logaritmo de las pernoctaciones hoteleras:

$$\hat{\log}(\text{turismo}) = 14.93182$$

El intervalo de predicción para el logaritmo de las pernoctaciones hoteleras se puede calcular también utilizando la información que da el output de Eviews:

$$\hat{V}(\hat{\log}(\text{turismo})) = 0.1209^2$$

$$s^2 = 0.4086^2$$

$$\hat{V}(e^0/X^P) = 0.4086^2 + 0.1209^2 = 0,18157$$

Entonces, el intervalo de confianza para la predicción de $\log(\text{turismo})$ al 95% es:

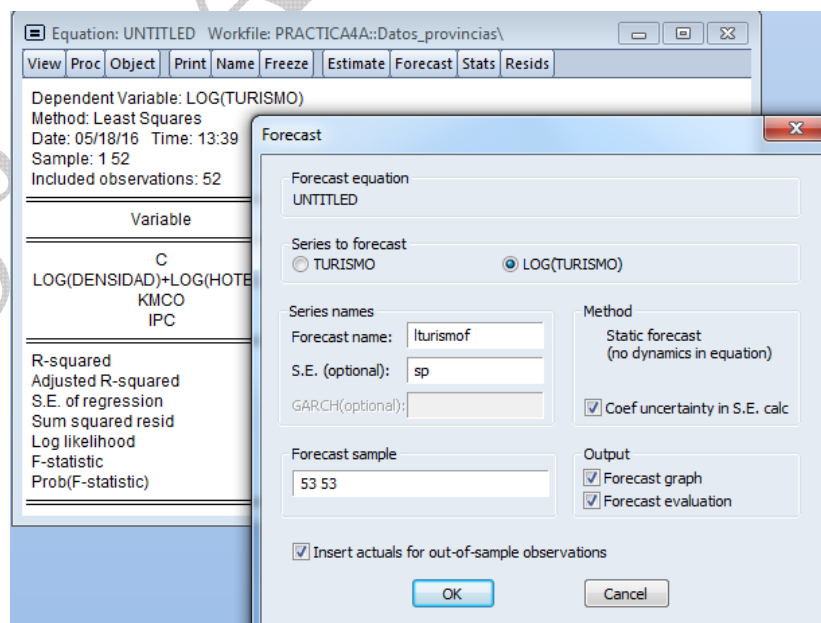
$$\hat{\log}(\text{turismo}) \pm t_{48;0,025} \cdot \sqrt{\hat{V}(e^0/X^P)}$$

$$14.93182 \pm 2.009 \cdot \sqrt{0.18157}$$

$$(14.076; 15.788)$$

La predicción se puede calcular también **automáticamente con Eviews**. Para ello, tenemos que:

- Expandir el rango del *workfile* para incluir una observación más: Range: 1 53
- Automáticamente se amplía también la muestra: Sample: 1 53
- Incluir en el *workfile* el valor de las explicativas para la observación 53.
- Estimar el modelo (observe que para la estimación Sample: 1 52)
- En el menú de la estimación, elegir la opción **Forecast**



- Podemos predecir las pernoctaciones hoteleras o el logaritmo de las pernoctaciones. En nuestro caso queremos predecir el logaritmo de las pernoctaciones, por lo que elegimos la opción "LOG(TURISMO)". Si eligiéramos la opción "TURISMO" se calcularía la predicción automáticamente con Eviews para el número de pernoctaciones hoteleras.
- Forecast name: nombre que se da a la predicción puntual ($\hat{\log}(\text{turismo})$). En nuestro caso la denominamos *lturismof*.
- S.E: nombre que se da a la desviación típica del error de predicción ($\sqrt{\hat{V}(e^0/X^P)}$). En nuestro caso la denominamos *sp*.
- Forecast sample: muestra para la que queremos predecir. En nuestro caso para la observación 53.

Una vez realizada la predicción aparecen dos nuevas series en el *workfile*, *lturismof* y *sp*:

- *lturismof* contiene la predicción puntual en la observación 53: $\hat{\log}(\text{turismo}) = 14.93182$. Esta predicción puntual coincide con la calculada anteriormente.
- *sp* contiene la desviación típica del error de predicción: $\sqrt{\hat{V}(e^0/X^P)} = 0.426193$

Con estos datos se puede construir un intervalo de confianza que coincide con el calculado anteriormente.

Predicción del número de pernoctaciones hoteleras: *turismo*

Respecto a la predicción puntual para la serie en nivel (número de pernoctaciones hoteleras), el valor *aproximado* de la predicción viene dado por la exponencial del valor predicho para la serie en logaritmos:

$$\hat{\text{turismo}} \approx e^{14.93182} = 3053564.0275$$

El intervalo de confianza para la predicción de la serie en nivel se calcula tomando la función exponencial para los valores extremos del intervalo de predicción calculado para la predicción de la serie en logaritmos. El intervalo de predicción para *turismo* al 95% es:

$$(e^{14.076}; e^{15.788}) = (1297565.014; 7188549.709)$$

Es el intervalo de posibles valores para el verdadero valor de la variable con un nivel de confianza del 95%. El intervalo de confianza es muy amplio por lo que la predicción realizada es muy poco precisa. Este resultado no es sorprendente dada la elevada dispersión que tienen algunas variables en la muestra considerada.

Para **predecir la serie en nivel automáticamente en Eviews**, procedemos de la misma manera que hemos explicado para la serie en logaritmos, pero ahora en el cuadro de diálogo de *Forecast*:

- Elegimos la opción "TURISMO".
- Forecast name: nombre de la predicción puntual ($turismo$). La denominamos *turismof*.
- S.E: no le damos nombre. El cálculo que hace Eviews en este caso no es correcto y no se debe utilizar para calcular el intervalo de confianza.
- Forecast sample: 53 53

Una vez realizada la predicción aparece la serie *turismof* en el *workfile*. *turismof* contiene la predicción puntual en la observación 53: $turismo = 3053577$. Esta predicción, al igual que la calculada anteriormente, es una *aproximación* de la verdadera predicción puntual.

El intervalo de confianza al 95% para la predicción de las pernoctaciones hoteleras es el que aparece en el gráfico de la predicción para la observación 53, delimitado por los puntos rojos (éstos coinciden con los extremos del intervalo calculado anteriormente).

