

**ECONOMETRIA I**  
**Grupo 4C – LADE**  
**Grupo piloto. Curso 2006 – 2007**  
**Profesor: Alfonso Novales**

**Hoja de ejercicios nº 2. Fecha de entrega: 31/10/2006**

(MC) En el archivo gta2.wf1 encontrará datos para 2005 correspondientes a las 12 sucursales que repartidas en 12 comunidades autónomas distintas tiene una determinada empresa, sobre las variables:

- \*  $ventas_i$  : ventas de la empresa en la CCAA  $i$ :
- \*  $precio_i$  : precio del producto que vende la empresa en la CCAA  $i$ :
- \*  $gpubli_i$  : gasto en publicidad de la empresa en la CCAA  $i$ :
- \*  $renta_i$  : renta media de la CCAA  $i$ :
- \*  $pcompe_i$  : precio de un producto competidor en la CCAA  $i$ :

Utilizando esta información, estime por mínimos cuadrados ordinarios los modelos:

$$ventas_i = \alpha + \beta precio_i + u_i \quad (M1)$$

$$ventas_i = \alpha + \beta gpubli_i + u_i \quad (M2)$$

$$ventas_i = \alpha + \beta_1 precio_i + \beta_2 gpubli_i + \beta_3 renta_i + \beta_4 pcompe_i + u_i \quad (M3)$$

y conserve los residuos que obtiene en cada uno de ellos.

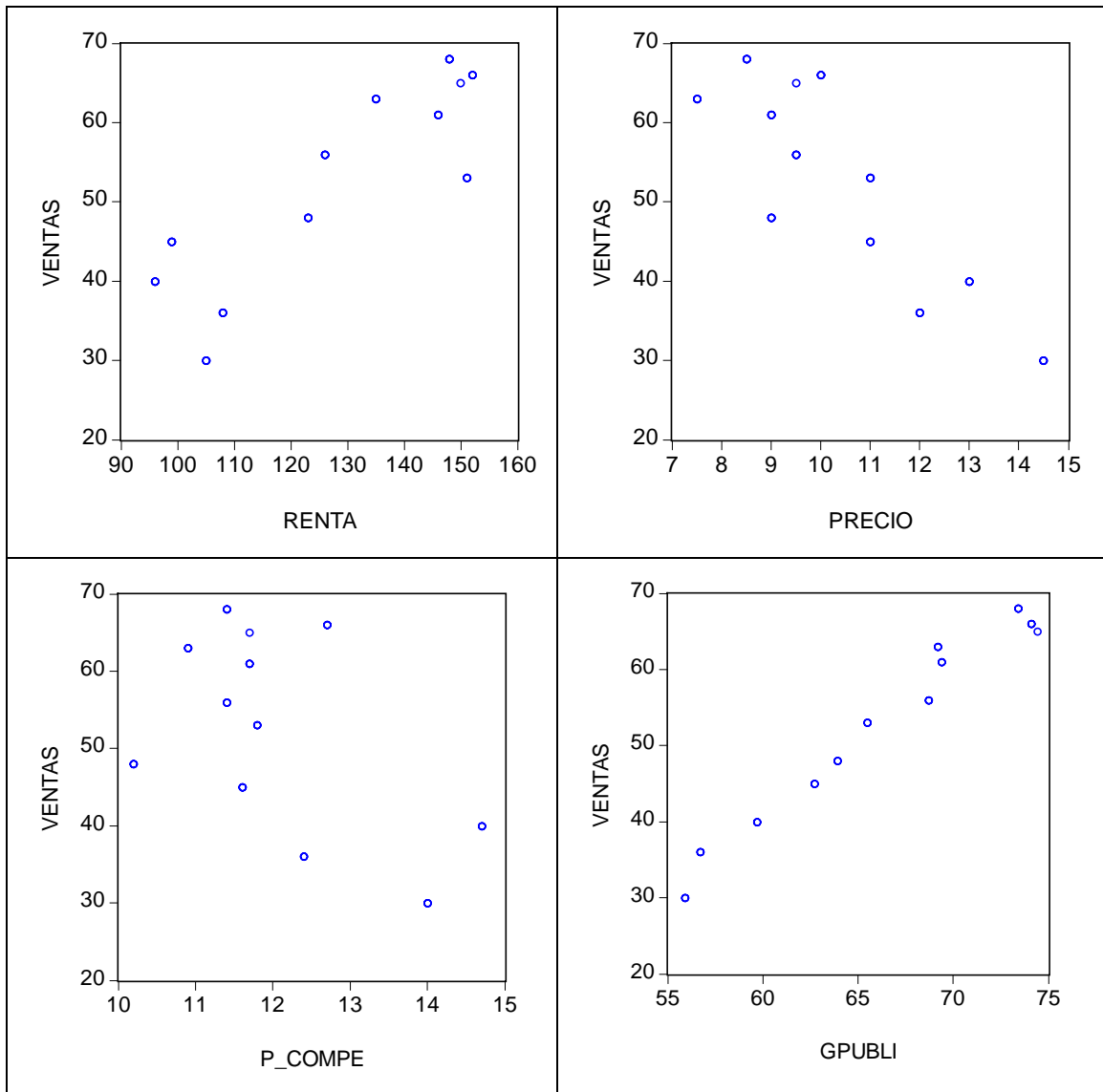
- a. Comente los resultados y analice el signo de los coeficientes.
- b. Compare los estadísticos R<sup>2</sup> de los tres modelos. ¿Son buenos modelos explicativos de las cifras de ventas?
- c. Examine los gráficos de ajuste y residuos de cada una de las 3 regresiones estimadas.
- d. Calcule el coeficiente de correlación entre los residuos de M1 y M3, así como entre M2 y M3. Haga un gráfico de nube de puntos para cada uno de estos dos pares de residuos. Comente sus resultados
- e. Tienen los tres conjuntos de residuos la misma media aritmética? ¿Y la misma desviación típica?
- f. Puede calcular los valores ajustados de cada regresión restando de la cifra de ventas los residuos ¿por qué? Calcule las medias muestrales de los valores ajustados que obtiene de cada una de las tres regresiones. ¿le sorprende el resultado que obtiene?
- g. Calcule las correlaciones lineales entre los residuos y la variable explicativa en cada regresión ¿le sorprende el resultado que obtiene? Calcule las correlaciones entre residuos y variable dependiente e interprete sus resultados.

**Solución del profesor:**

**a), b), c) Comente resultados de regresiones, coeficientes R<sup>2</sup>, gráficos de ajuste y residuos**

Comenzamos examinando la información disponible. Disponemos de información acerca de 12 Comunidades Autónomas. Es, por tanto, una muestra de corte transversal. Hemos de tener cuidado en no interpretar los gráficos que hagamos como si de una evolución temporal se tratase. De hecho, esta vez no vamos a construir los habituales gráficos de líneas para evitar este posible error de interpretación. En su lugar,

presentamos nubes de puntos de la variable dependiente (Ventas), con cada una de las posibles variables explicativas. Es preferible tener la variable dependiente en ordenadas, lo que se consigue marcando primero el nombre de la variable explicativa (precio, por ejemplo) y luego el de las Ventas.



Como puede apreciarse, las ventas guardan una relación positiva con la renta de la CCAA y también con el gasto en publicidad en dicha CCAA. La relación con el gasto en publicidad parece todavía más estrecha que con la renta. Por otro lado, las ventas parecen estar negativamente relacionadas con el precio del producto, pero también con el precio de la competencia. Esta última relación es algo sorprendente, si bien no parece una relación muy ajustada. Podría suceder que nuestro precio y el de la competencia estén relacionados por razones de funcionamiento del mercado, y que ninguno de ellos explique mucho la cifra de ventas.

La matriz de coeficientes de correlación sugiere que el precio fijado por la competencia es menos importante que las restantes variables para explicar las ventas.

	VEN	REN	GPUB	PREC	P_COM
VENTAS	1.00				
RENTA	0.86	1.00			
GPUBLI	0.98	0.85	1.00		
PRECIO	-0.85	-0.66	-0.79	1.00	
P_COMPE	-0.53	-0.46	-0.48	0.83	1.00

Estimamos las tres regresiones que se nos pide, obteniendo,

### Modelo 1:

Dependent Variable: VENTAS		Method: Least Squares		
Sample: 1 12		Included observations: 12		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	108.6054	11.01569	9.859153	0.0000
PRECIO	-5.399719	1.043899	-5.172643	0.0004
R-squared	0.727937	Mean dependent var	52.58333	
Adjusted R-squared	0.700731	S.D. dependent var	12.73833	
S.E. of regression	6.968566	Akaike info criterion	6.871708	
Sum squared resid	485.6092	Schwarz criterion	6.952526	
Log likelihood	-39.23025	F-statistic	26.75624	
Durbin-Watson stat	0.592157	Prob(F-statistic)	0.000418	

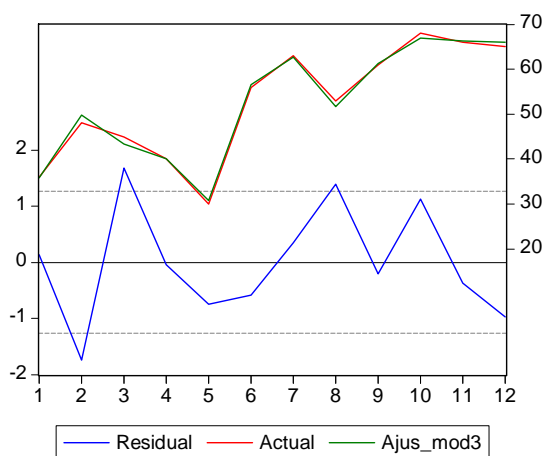
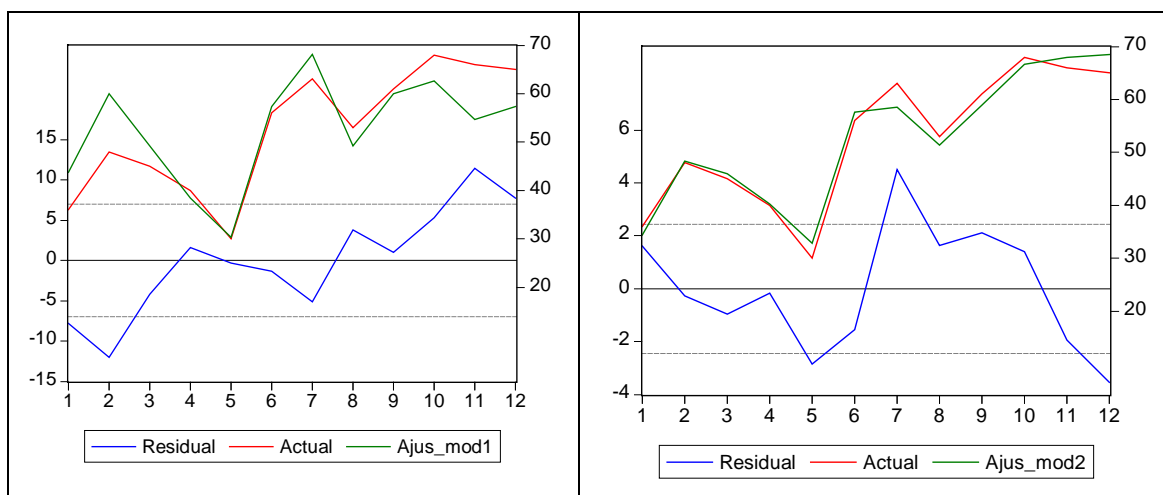
### Modelo 2:

Dependent Variable: VENTAS		Method: Least Squares		
Sample: 1 12		Included observations: 12		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-75.00002	7.539007	-9.948262	0.0000
GPUBLI	1.929184	0.113497	16.99767	0.0000
R-squared	0.966546	Mean dependent var	52.58333	
Adjusted R-squared	0.963201	S.D. dependent var	12.73833	
S.E. of regression	2.443605	Akaike info criterion	4.775837	
Sum squared resid	59.71204	Schwarz criterion	4.856655	
Log likelihood	-26.65502	F-statistic	288.9207	
Durbin-Watson stat	1.215309	Prob(F-statistic)	0.000000	

### Modelo 3:

Dependent Variable: VENTAS		Method: Least Squares		
Sample: 1 12		Included observations: 12		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-30.55653	9.549212	-3.199901	0.0151
PRECIO	-2.842484	0.597282	-4.759031	0.0021
GPUBLI	1.167260	0.162942	7.163640	0.0002
RENTA	0.088978	0.034783	2.558074	0.0377
P_COMPE	1.995106	0.672916	2.964868	0.0210
R-squared	0.993693	Mean dependent var	52.58333	
Adjusted R-squared	0.990089	S.D. dependent var	12.73833	
S.E. of regression	1.268128	Akaike info criterion	3.607298	
Sum squared resid	11.25705	Schwarz criterion	3.809342	
Log likelihood	-16.64379	F-statistic	275.7299	
Durbin-Watson stat	2.458923	Prob(F-statistic)	0.000000	

Los tres modelos presentan una capacidad explicativa apreciable, aunque vemos que el gasto en publicidad parece explicar las ventas mejor que el precio del bien., con un R2 de 0,97 en el primer caso, frente a un R2 de 0,73 del segundo modelo. El R2 del tercer modelo es aun más alto, como debe ser, por incluir a ambas variables como explicativas. Para comparar su ajuste con el de los Modelos 1 y 2, debemos examinar el R2 corregido, que es de 0,99. En definitiva, las diferencias existentes en las cuatro variables que aparecen como explicativas en este modelo permiten explicar la mayor parte de la variación en ventas a lo largo de las distintas CCAA.



Para percibir mejor los diferentes ajustes de cada modelo, podemos examinar los gráficos de ajuste y residuos que muestran un ajuste relativamente bueno del Modelo 1, un ajuste aún mejor del Modelo 2, y un ajuste casi perfecto del Modelo 3.

En cuanto a los signos de los coeficientes estimados, es razonable el signo negativo del precio propio sobre las ventas, así como los signos positivos del efecto sobre las ventas de variaciones en renta, el gasto en publicidad y el precio de la competencia.

Es interesante apreciar que en la regresión que incorpora las cuatro variables explicativas, el precio de la competencia aparece afectando a las ventas con el signo positivo que cabría esperar.

#### d) Correlaciones entre residuos de distintos modelos

Los coeficientes de correlación entre los residuos de los distintos modelos, así como entre cada uno de estos y la variable dependiente Ventas son,

	VENTAS	RES_MOD1	RES_MOD2	RES_MOD3
VENTAS	1.00			
RES_MOD1	0.52	1.00		
RES_MOD2	0.18	-0.35	1.00	
RES_MOD3	0.08	0.15	0.43	1.00

que sugiere que los residuos de los tres modelos son distintos entre sí. Ello puede interpretarse como evidencia acerca de que el Modelo 3 es significativamente mejor que los otros dos, así como que el Modelo 2 proporciona una explicación significativamente distinta (y mejor) que el Modelo 1. Por otra parte los tres conjuntos de residuos apenas muestran correlación con las Ventas, indicando que cada uno de los tres modelos tiene una cierta capacidad explicativa. Nótese que si un modelo que pretende explicar las Ventas no tiene apenas capacidad explicativa, entonces la variable dependiente del mismo (Ventas) y los residuos serán muy similares entre sí.

e) Los estadísticos descriptivos de los residuos de los 3 modelos estimados muestran medias aritméticas iguales a cero, si bien valores mediana no nulos. La desviación típica del residuos del Modelo1 es mayor que la del residuo del Modelo2 y la de éste es mayor que la del Modelo 3, como corresponde a la distinta capacidad explicativa de cada modelo, a la que ya hemos hecho referencia. Cuanto mayor sea la capacidad explicativa de un modelo, menor será la desviación típica de su residuo. No se rechaza la hipótesis nula de Normalidad en ninguno de los tres residuos, pues el valor-p del contraste de Jarque-Bera está muy por encima de los niveles de significación habituales.

	RES_MOD1	RES_MOD2	RES_MOD3
Mean	-2.54E-14	2.09E-14	1.08E-14
Median	0.341281	-0.223539	-0.124527
Maximum	11.39177	4.500509	1.684590
Minimum	-12.00795	-3.531255	-1.743415
Std. Dev.	6.644267	2.329886	1.011616
Skewness	-0.089500	0.247861	0.186817
Kurtosis	2.383869	2.321197	2.274337
Jarque-Bera	0.205829	0.353257	0.333095
Probability	0.902204	0.838091	0.846583
Sum	-3.07E-13	2.50E-13	1.30E-13
Sum Sq. Dev.	485.6092	59.71204	11.25705
Observations	12	12	12

f) Las observaciones muestrales de la variable dependiente pueden descomponerse,

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i = \hat{y}_i + \hat{u}_i$$

donde  $y$  representa la variable Ventas. Aunque la descomposición anterior describe un modelo con una sola variable explicativa, la extensión al caso en que se utilice más de una de dichas variables es evidente. Por tanto, podemos generar en EVIEWS los valores ajustado por un modelo sin más que restar de las observaciones de la variable dependiente, los residuos del modelo. Esto es lo que hemos hecho en el archivo adjunto.

Si tomamos promedio a lo largo de la ecuación anterior, y puesto que la media de los residuos es cero, tendremos que, al estimar por mínimos cuadrados, la media muestral de los valores ajustados por el modelo coincide con la media muestral de la variable dependiente. Por tanto, si, como es el caso en nuestro ejercicio, tenemos tres modelos alternativos para explicar la misma variable dependiente, los valores ajustados por los tres modelos tendrán la misma media muestral, la de la variable Ventas. La siguiente tabla de estadísticos descriptivos muestra que, en efecto, las cifras de Ventas explicadas por los tres modelos tienen la misma media muestral.

	EXPLICADO _MOD1	EXPLICADO _MOD2	EXPLICADO _MOD3
Mean	52.58333	52.58333	52.58333

Median	55.95816	54.44821	54.09785
Maximum	68.10753	68.53126	66.87221
Minimum	30.30949	32.84135	30.75147
Std. Dev.	10.86825	12.52345	12.69810
Skewness	-0.653389	-0.241007	-0.380241
Kurtosis	2.618439	1.814698	1.768082
Jarque-Bera	0.926629	0.818639	1.047978
Probability	0.629195	0.664102	0.592154
Sum	631.0000	631.0000	631.0000
Sum Sq. Dev.	1299.308	1725.205	1773.660
Observations	12	12	12

- g) Los coeficientes de correlación entre los residuos de cada modelo y las variables explicativas del mismo serán cero, por la propiedad del estimador de mínimos cuadrados. No las incluimos aquí, aunque el lector puede calcularlas fácilmente y comprobar esta afirmación.

Por otro lado, y como ya hemos comentado más arriba, debemos esperar que cuanto mejor sea el modelo econométrico, mayor sea la correlación de la variable dependiente con los valores ajustados y menor sea la correlación con los residuos.

	VENTAS	EXPLICADO_MOD1	EXPLICADO_MOD2	EXPLICADO_MOD3
VENTAS	1.00			
EXPLICADO_MOD1	0.85	1.00		
EXPLICADO_MOD2	0.98	0.79	1.00	
EXPLICADO_MOD3	0.99	0.85	0.98	1.00

La tabla de correlaciones entre los residuos de los tres modelos y la variable Ventas que antes vimos, así como la tabla de correlaciones entre las Ventas y los valores ajustados por los tres modelos que acabamos de proporcionar son, por tanto, totalmente coherentes con la información proporcionada tanto por los coeficientes R2 como por las desviaciones típicas de los residuos de los tres modelos. El lector debe asegurarse de que entiende esta coherencia entre los cuatro indicadores mencionados: correlación entre variable dependiente y valores ajustados, correlación entre variable dependiente y residuos, R2 y desviación típica de los residuos.