Taller 4: Regresión lineal múltiple
Econometría 06216
08-23-2010

Profesor: Julio César Alonso C
Monitores: Sasha Magyaroff - Carolina Restrepo.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 30 de agosto de 2010. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**.

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo individual. Por tanto el taller debe reflejar únicamente el trabajo del estudiante.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.

Papá pitufo desea realizar un estudio sobre el comportamiento del precio de las casas en su Aldea, para ello cuenta con los datos del archivo T4-02-10.XLS que contiene la siguiente información: Y_t es el precio en dólares de las casas, x_{2t} es el tamaño de las casas en metros cuadrados y x_{3t} es el número de habitaciones de cada casa.

Papa pitufo basado en una propuesta de Pitufina decide estimar el siguiente modelo:

$$Y_i = (Sen^2\beta_5 + Cos^2\beta_6)^{\beta_3} x_{2i}^{\beta_2} x_{3i}^{\beta_3} \varepsilon_i{}^{\beta_5} \quad (1)$$

Pregunta 1

a) Pruebe que el modelo puede estimarse por MCO. De no ser posible explique la razón del por qué no se puede estimar.
b) Interprete los coeficientes a priori, y justifique el signo esperado.

El pitufo Gruñón advierte que para explicar el comportamiento de los precios se podrían estimar los siguientes modelos alternativos: El primero de ellos tiene todas las variables en nivel:

$$Y_i = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \mu_i \quad (2)$$

El segundo modelo usa logaritmo en todas las variables excepto en x_{3t}.y agrega una variable Dummy que toma el valor de 1 si la casa está en condominio cerrado y 0 en caso contrario.

$$lnY_i = \gamma_1 + \gamma_2 lnx_{2i} + \gamma_3 x_{3i} + \delta_1 D_i + \mu_i \tag{3}$$

Donde, $D_i = \begin{cases} 1 & La\ vivienda\ se\ encuentra\ en\ condominio\ cerrado \\ 0 & o.w \end{cases}$

El tercer modelo usa logaritmos en todas las variables y agrega una variable Dummy que toma el valor de 1 si la casa está en condominio cerrado y 0 en caso contrario.

$$lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \delta_1 D_i + \mu_i \tag{4}$$

Donde, $D_i = \begin{cases} 1 & La\ vivienda\ se\ encuentra\ en\ condominio\ cerrado \\ 0 & o.w \end{cases}$

Pregunta 2

a) Estime los modelos de las ecuaciones (1), (2),(3) y (4) y preséntelos en una tabla (Tabla 1).
b) A partir de sus resultados: Interprete los coeficientes del modelo (4) teniendo en cuenta su significancia.

Pregunta 3

Continuando con la pregunta anterior.
a) ¿Cuál de los cuatro modelos es mejor? Sea lo más claro posible en su respuetas.
b) Elabore las tablas ANOVA de los modelos (comparables) e indique cuáles de las celdas de las respectivas tablas permanecen constantes y cuáles cambian ante cambios en la especificación en el modelo que impliquen: i). Cambios en la variable dependiente solamente; ii). Cambios en la especificación de las variables independientes solamente, iii).Cambios en el número de variables independientes solamente.

Pregunta 4

Por otro lado, el Alcalde de la Aldea el Dr. Gargamel afirma que el mejor modelo es el de la (1). Además, algunos vecinos de la Aldea han escuchado al Dr. Gargamel afirma públicamente que:

$$H_0: \beta_2 + \beta_3 = 1$$
$$H_1: No\ H_0$$

El pitufo miedoso, ignorando los resultados de los puntos anteriores es el único pitufo que le hace caso al Dr. Gargamel y prueba la restricción (con la especificación (1)).
a) Escriba la restricción del tipo R β =C y encuentre el valor del F-calculado que corresponde a ésta prueba de hipótesis.

b) ¿Es correcta la afirmación pública del alcalde sobre la suma de los coeficientes?

Pregunta 5

Algunos vecinos de la Aldea han escuchado al Dr. Gargamel afirmar públicamente que los precios de las casas se comportan totalmente diferentes de acuerdo a la zona de la ciudad. Concretamente el pitufo Filósofo lo escuchó decir que los precios se ven afectados por la localización geográfica (sur-norte) debido a que se valoriza más en el sur.

a) Escriba el modelo que debe ser estimado (a partir del modelo (1)).

b) Demuestre, antes de estimar, qué el modelo que escribió en la pregunta anterior es el adecuado para determinar si el alcalde tiene o no la razón sobre el efecto de la localización sobre el precio.

Pregunta 6

a). Estime el modelo que construyó en la pregunta anterior y repórtelo en una nueva tabla (tabla 2)

b). Comente la significancia individual de los coeficientes estimados.

c). ¿Tiene el alcalde la razón en su afirmación? Sea lo más claro posible y muestre todos sus cálculos.

d). Interprete el R^2 del modelo estimado.

e). Compare este modelo con el elegido en la pregunta 3: ¿Cuál es mejor?

Taller 4: Regresión lineal múltiple
Econometría 06216
08-23-2010

Profesor: Julio Cesar Alonso
Monitores: Sasha Magyaroff – Carolina Restrepo

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas
- Este taller deberá subirse a la plataforma Moodle hasta las 7:10 del 30 de agosto de 2010. **Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta**.

Instrucciones:

- Este taller debe ser escrito en computador. Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo individual. Por tanto el taller debe reflejar únicamente el trabajo del estudiante.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.

Papá pitufo desea realizar un estudio sobre el comportamiento del precio de las casas en su Aldea, para ello cuenta con los datos del archivo T4-02-10.XLS que contiene la siguiente información: Y_i es el precio en dólares de las casas, x_{2i} es el tamaño de las casas en metros cuadrados y x_{3i} es el número de habitaciones de cada casa.

Papa pitufo basado en una propuesta de Pitufina decide estimar el siguiente modelo:

$$Y_i = (Sen^2\beta_5 + Cos^2\beta_6)^{\beta_3} x_{2i}^{\beta_2} x_{3i}^{\beta_3} \varepsilon_i{}^{\beta_5} \qquad (1)$$

Pregunta 1

a) Pruebe que el modelo puede estimarse por MCO. De no ser posible explique la razón del por qué no se puede estimar.

$$lnY_i = \beta_3 ln(Sen^2\beta_5 + Cos^2\beta_6) + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \beta_5 ln\epsilon_i$$

$$lnY_i = \beta_3 ln(Sen^2\beta_5 + Cos^2\beta_6) + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \beta_4 ln\epsilon_i$$

$$lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \mu_t$$

Por lo tanto el modelo es linealizable y puede ser estimado por el método de los MCO.

b) Interprete los coeficientes a priori, y justifique el signo esperado.

β_1 No tiene interpretación económica.

β_2 Es la elasticidad del precio de la vivienda con respecto al tamaño de la casa y se espera que su signo sea positivo. Un incremento de 1% en los metros cuadrados del tamaño de la casa, aumenta en promedio β_2 % el precio de la vivienda.

β_3 Es la elasticidad del precio con respecto al número de habitaciones y se espera que su signo sea positivo. Un incremento de 1% en el número de habitaciones generaría, en promedio y ceteris paribus, un incremento del β_3 % en el precio de la vivienda.

El pitufo Gruñón advierte que para explicar el comportamiento de los precios se podrían estimar los siguientes modelos alternativos: El primero de ellos tiene todas las variables en nivel:

$$Y_i = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \mu_i \qquad (2)$$

El segundo modelo usa logaritmo en todas las variables excepto en x_{3t}.y agrega una variable Dummy que toma el valor de 1 si la casa está en condominio cerrado y 0 en caso contrario.

$$lnY_i = \gamma_1 + \gamma_2 lnx_{2i} + \gamma_3 x_{3i} + \delta_1 D_i + \mu_i \qquad (3)$$

Donde, $D_i = \begin{cases} 1 & La\ vivienda\ se\ encuentra\ en\ condominio\ cerrado \\ 0 & o.w \end{cases}$

El tercer modelo usa logaritmos en todas las variables y agrega una variable Dummy que toma el valor de 1 si la casa está en condominio cerrado y 0 en caso contrario.

$$lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \delta_1 D_i + \mu_i \qquad (4)$$

Donde, $D_i = \begin{cases} 1 & La\ vivienda\ se\ encuentra\ en\ condominio\ cerrado \\ 0 & o.w \end{cases}$

Pregunta 2

a) Estime los modelos de las ecuaciones (1), (2),(3) y (4) y preséntelos en una tabla (Tabla 1).

Tabla 1. Estimaciones del modelo 1, modelo 2, modelo 3, modelo 4

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	M.C.O	M.C.O	M.C.O	M.C.O
Constante	-0.78819 (-1.146)	-19.37711 (-0.624)	-0,67752 (-0.970)	-0.84654 (-1.232)
X2		0.12844*** (9.290)		
X3		15.21379 (1.604)	0.02790 (0.880)	
LnX2	0.83376 (8.435)***		0.81523*** (8.253)	0.84374 (8.533)***
LnX3	0.08640 (0.768)			0.03816 (0.321)
D			0.05443 (1.101)	0.06187 (1.231)
R^2	0.5561	0.6319	0.5674	0.5639

R^2-Ajustado	0.5456	0.6233	0.5519	0.5484
F	53.24	72.9721	36.72	36.21
N° de Obs.	88	88	88	88

* Nivel de significancia 90%
** Nivel de significancia 95%
*** Nivel de significancia 99%
MCO: Mínimos Cuadrados Ordinarios.

b) A partir de sus resultados: Interprete los coeficientes del modelo 4 teniendo en cuenta su significancia.

$\hat{\beta}_1 = -0.84654.$ No es significativo, es decir, es estadísticamente igual a cero para el modelo.

$\hat{\beta}_2$ =0.84374. Este parámetro es significativo al 99%: Un incremento del 1% en los metros cuadrados de la vivienda aumentará en promedio en 0,84374 por ciento el precio de la vivienda.

$\hat{\beta}_3 = 0.03816.$ Este coeficiente no es significativo para el modelo.

$\hat{\delta}_1 = 0.06187.$ Este coeficiente tampoco es significativo en el modelo.

Pregunta 3
Continuando con la pregunta anterior.

a) ¿Cuál de los cuatro modelos es mejor? Sea lo más claro posible en sus respuestas.

Noten que no todos los modelos (1,2,3 y 4) se pueden comparar debido a que el modelo 2 no tiene la misma variable dependiente que los modelos 1, 3 y 4. Los modelos 1, 3 y 4 si se pueden comparar y se hace con el R2-ajustado pues no todos tienen el mismo número de k y el R2 puede estar sobre estimado. El modelo con el mejor R2-ajustado es el modelo de la ecuación 3 con un R2-ajustado 55.19%, mostrando la mejor bondad de ajuste de los tres.

b) Elabore las tablas ANOVA de los modelos (comparables) e indique cuáles de las celdas de las respectivas tablas permanecen constantes y cuales cambian ante cambios en la especificación en el modelo que impliquen: i). Cambios en la variable dependiente solamente; ii). Cambios en la especificación de las variables independientes solamente, iii).Cambios en el numero de variables independientes solamente.

Tabla ANOVA ecuación 1

	SS	**GL**	**MS**
SSR	4.45	2	2.22
SSE	3.55	85	0.04
SST	8.01	87	0.09

Tabla ANOVA ecuación 3

	SS	**GL**	**MS**
SSR	4.54	3	1.51
SSE	3.46	84	0.04
SST	8.01	87	0.09

Tabla ANOVA ecuación 4

	SS	GL	MS
SSR	4.52	3	1.50
SSE	3.49	84	0.04
SST	8.01	87	0.09

Las entradas de la Tabla ANOVA que permanecerían inalteradas ante cambios en la especificación de los modelos con una misma variable explicada pero con un número igual de variables explicativas serían: las dos que corresponden a los grados de libertad de la regresión y el error de la ecuación 3 y 4, así como la suma de cuadrados total.

Las entradas que permanecerían inalteradas ante cambios en la especificación de las variables independientes solamente, serían los grados de libertad y la suma de cuadrados total.

La única entrada que permanecería inalterada ante cambios en el número de variables independientes solamente sería SST.

Pregunta 4

Por otro lado, el Alcalde de la Aldea el Dr. Gargamel afirma que el mejor modelo es el (1). Además algunos vecinos de la Aldea han escuchado al Dr. Gargamel afirma públicamente que: $H_0: \beta_2 + \beta_3 = 1$ y por tanto que $H_1: No\ H_0$. El pitufo miedoso, ignorando los resultados de los puntos anteriores es el único pitufo que le hace caso al Dr. Gargamel y prueba la restricción (con la especificación 1).

a) Escriba la restricción del tipo $RB = C$ y encuentre el valor del F-calculado que corresponde a esta prueba de hipotesis.

Las hipótesis a probar son: $H_0: \beta_2 + \beta_3 = 1$ vs $H_1: No\ H_0$. Esta hipótesis la podemos probar de la forma $R_{r*k}\beta_{k*1} = C_{r*1}$; mediante una prueba F. O si lo hacemos en EasyReg debemos usar el Test de Wald. Utilizando el EasyReg se encuentra que el estadístico de wald es 2.61, y el p valor correspondientes es 0.10602. Por lo tanto, bajo un nivel de significancia del 90%, no existe evidencia suficiente para rechazar la idea de que la suma de los coeficientes asociados a las variables tamaño y número de habitaciones es igual a uno.

b) ¿Es correcta la afirmación pública del alcalde sobre la suma de los coeficientes?.

Como se mencionó en el numeral anterior, no existe evidencia suficiente para probar que la suma de los coeficientes $\beta_2\ y\ \beta_3$ es diferente de 1. Suponiendo que la especificación del modelo 1 es la adecuada, la afirmación pública del alcalde es en promedio correcta.

Pregunta 5

Algunos vecinos de la Aldea han escuchado al Dr. Gargamel afirmar públicamente que los precios de las casas se comportan totalmente diferente de acuerdo a la zona de la ciudad. Concretamente el pitufo Filósofo lo escuchó decir que los precios se ven afectados por la localización geográfica (sur-norte) debido a que se valoriza más en el sur.

a) Escriba el modelo que debe ser estimado (a partir del modelo 1).

A partir del modelo (1) $lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \mu_i$, se puede reescribir el modelo que recoja la nueva variable explicativa así:

$$lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \alpha_1 D_i + \alpha_2 D_i lnx_{2i} + \alpha_3 D_i lnx_{3i} + \mu_t \quad \text{Modelo (5)}$$

Donde

$$D_i = \begin{cases} 1 & \text{Si la casa se encuentra en el sur} \\ 0 & \text{Si la casa se encuentra en el Norte} \end{cases}$$

b) Demuestre antes de estimar, que el modelo que escribió en la pregunta anterior es el adecuado para determinar si el alcalde tiene o no la razón sobre el efecto de la localización sobre el precio.

Este análisis implica determinar el valor esperado del modelo que se plantea en el numeral a:

$$E[lnY_i] = \begin{cases} (\beta_1 + \alpha_1) + lnx_{2i}(\beta_2 + \alpha_2) + lnx_{3i}(\beta_3 + \alpha_3) & \text{Si se encuentra en el sur} \\ \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} & \text{Si se encuentra en el Norte} \end{cases}$$

De esta manera si la estructura de precios se ve afectada por la localización geográfica, entonces todas las variables del modelo se ven afectadas, por lo tanto la variable Dummy tiene un efecto sobre todo el modelo. Por otra parte si no existen diferencias en la estructura de precios entre estar en el sur o en el norte, entonces el modelo (5) colapsa en el modelo 1.

Pregunta 6

a. Estime el modelo que construyó en la pregunta anterior y repórtelo en una nueva tabla (tabla 2)

Tabla 2. Estimación del Modelo 5

Variable dependiente Ln Y M.C.O *t estadístico entre paréntesis*	Modelo 5	Modelo 6
Constante	-0.22211	-0.7902336
	(-0.227)	(-1.151)
LnX2	0.72513***	0.8341119***
	(-5.277)	(8.452)
LnX3	0.30478**	0.1030967
	(-2.127)	(0.910)
D_i	-1.42759	-
	(-1.050)	
D_iLnX2	0.27248	
	(-1.391)	-
D_iLnX3	-0.53806**	-0.0389273
	(-2.395)	(-1.124)
R^2	0.5883	0.5627
R^2-Ajustado	0.5632	0.547
F	23.44 ***	36.02
Nº de Obs.	88	88

* Nivel de significancia 90%
** Nivel de significancia 95%

*** Nivel de significancia 99%

b. Comente la significancia individual de los coeficientes estimados.
Para la especificación

$$lnY_i = \beta_1 + \beta_2 lnx_{2i} + \beta_3 lnx_{3i} + \alpha_1 D_i + \alpha_2 D_i lnx_{2i} + \alpha_3 D_i lnx_{3i} + \mu_t$$

Se encontró que β_2 es significativo al 99%, β_3 Es significativo al 95% y el coeficiente de $D_i LnX3$ Es significativo al 95%, las de mas variables no son significativas ni al 90%,95% o 99%.

c. ¿Tiene el alcalde la razón en su afirmación?

Establecer si el alcalde tenía razón o no, nos remite a testear si los coeficientes asociados a las variables Dummy son o no estadísticamente significativos. De ser significativos, se encontraría que el alcalde está en lo cierto. El estadístico Wald para probar la H0: $\alpha_1 = \alpha_2 = \alpha_3$ =0 versus la hipótesis alterna HA: No H0 es 6.42. Este estadístico permite rechazar la hipótesis nula únicamente a un nivel de significancia del 10%.

Adicionalmente, dado que sólo uno de los coeficientes asociados a la variable dummy (que recoge la diferencia en los precios de una casa derivados de su ubicación) es significativo (α_3), para determinar si el alcalde tiene razón en su afirmación sería conveniente realizar una última estimación con un intercepto γ_1 que considere los resultados de significancia de la estimación anterior así:

$$lnY_i = \gamma_1 + \gamma_2 lnx_{2i} + \gamma_3 lnx_{3i} + \gamma_4 D_i lnx_{3i} + \varepsilon_i \quad \text{Modelo (6)}$$

$$\text{donde, } D_i = \begin{cases} 1 & Si\ la\ casa\ se\ encuentra\ en\ el\ sur \\ 0 & Si\ la\ casa\ se\ encuentra\ en\ el\ Norte \end{cases}$$

La estimación de este modelo se presenta en la Tabla 2. Como puede observarse, la prueba de significancia individual del coeficiente γ_4 permite concluir al 90%, 95% y 99% que dicho coeficiente es significativamente igual a cero. En otras palabras, que el alcalde NO tiene razón en su afirmación.

d. Interprete el R^2 del modelo estimado.
El 58,83% de la variabilidad del logaritmo natural del precio de la vivienda es explicado por el modelo.

d. Compare este modelo con el elegido en la pregunta 3: ¿Cuál es mejor?
Con respecto al tercer modelo, como ambos presentan la misma variable dependiente, podemos comparar el R^2 ajustado y concluir que el modelo (5) presenta un mejor ajuste.

Precio	No Habi	Tamaño	condominio	Localizacion
300	4	2438	1	0
370	3	2076	1	1
191	3	1374	0	1
195	3	1448	1	0
373	4	2514	1	1
467	5	2754	1	0
332,5	3	2067	1	1
315	3	1731	1	0
206	3	1767	0	1
240	3	1890	0	0
285	4	2336	1	1
300	5	2634	1	0
405	3	3375	1	1
212	3	1899	0	1
265	3	2312	1	1
227,4	4	1760	1	1
240	4	2000	0	0
285	3	1774	1	0
268	3	1376	1	0
310	4	1835	1	1
266	3	2048	1	1
270	3	2124	1	1
225	3	1768	0	0
150	4	1732	0	1
247	3	1440	1	0
275	3	1932	0	0
230	3	1932	0	1
343	3	2106	1	0
477,5	7	3529	1	1
350	4	2051	1	0
230	4	1573	1	1
335	4	2829	0	0
251	3	1630	1	0
235	4	1840	1	1
361	4	2066	1	1
190	4	1702	0	1
360	4	2750	1	0
575	5	3880	1	1
209	4	1854	1	0
225	2	1421	0	0
246	3	1662	1	0
713,5	5	3331	1	0
248	4	1656	1	1
230	3	1171	0	1
375	5	2293	1	0
265	3	1764	1	0

313	3	2768	0	0
417,5	4	3733	0	0
253	3	1536	1	1
315	4	1638	1	0
264	3	1972	1	1
255	2	1478	0	1
210	3	1408	1	0
180	3	1812	1	0
250	3	1722	1	0
250	4	1780	1	0
209	4	1674	1	1
258	4	1850	1	0
289	3	1925	1	0
316	4	2343	0	0
225	3	1567	0	1
266	4	1664	1	0
310	6	1386	1	0
471,25	5	2617	1	0
335	4	2321	1	0
495	4	2638	1	0
279,5	4	1915	1	1
380	4	2589	1	1
325	4	2709	0	1
220	3	1587	1	0
215	3	1694	0	0
240	3	1536	1	1
725	5	3662	0	1
230	3	1736	1	0
306	2	2205	0	0
425	3	1502	0	1
318	4	1696	1	0
330	3	2186	1	0
246	4	1928	1	0
225	3	1294	0	1
111	4	1535	1	1
268	3	1980	1	0
244	4	2090	1	1
295	3	1837	1	1
236	3	1715	0	0
202,5	3	1574	0	0
219	2	1185	0	1
242	4	1774	1	0