

Taller 3: Regresión Múltiple

Econometría 06216

08-02-2010

Profesores: Julio César Alonso.

Monitoras: 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 AM del 16 de Agosto de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

- 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.
- Este taller debe ser escrito en computador.

Pregunta 1

Para las preguntas 1 a 3, realice los cálculos de manera manual. Presente todo el procedimiento y aclare muy bien cuáles fórmulas fueron empleadas y qué valores fueron reemplazados

Una reconocida distribuidora de elementos deportivos (patines, balones, bicicletas, accesorios, entre otros), supone que las ganancias de las ventas y_t (medidas en pesos) de sus productos sigue la siguiente relación:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \varepsilon_t \quad (1)$$

Donde X_{2t} corresponde a la cantidad de patines vendidos en el último mes; X_{3t} es el número de devoluciones mensuales por daño o imperfección; X_{4t} es el logaritmo del porcentaje de descuento ofrecido en el mes t . Además, ε_t representa una perturbación aleatoria que cumple los supuestos del teorema del Gauss-Markov.

Para 1000 observaciones se obtuvieron los siguientes valores:

$$\begin{array}{cccc}
 \sum_{t=1}^n X_{2t} = 4000 & \sum_{t=1}^n X_{3t} = 500 & \sum_{t=1}^n X_{4t} = 1917 & \\
 \sum_{t=1}^n (X_{2t})^2 = 10000 & \sum_{t=1}^n (X_{3t})^2 = 1000 & \sum_{t=1}^n (x_{4t})^2 = 4000 & \sum_{t=1}^n y_t = 5200 \\
 \sum_{t=1}^n X_{2t} \cdot X_{3t} = 2000 & \sum_{t=1}^n X_{2t} \cdot X_{4t} = 3200 & \sum_{t=1}^n X_{3t} \cdot X_{4t} = 800 & \sum_{t=1}^n y_t \cdot X_{3t} = 6400 \\
 \sum_{t=1}^n y_t \cdot X_{2t} = 11000 & \sum_{t=1}^n (y_t)^2 = 57098.36 & \sum_{t=1}^n y_t \cdot X_{4t} = 10500 &
 \end{array}$$

Construya las matrices: $X^T X$, $X^T y$, $y^T y$

Pregunta 2

Continuando con el enunciado de la pregunta anterior, estime el modelo 1.

Pregunta 3

Continuando con la pregunta anterior, responda:

- a. Interprete los coeficientes estimados en la pregunta 2.
- b. Discuta la significancia conjunta de los coeficientes estimados

Pregunta 4

El gobierno de una pequeña República está reconsiderando la viabilidad del transporte ferroviario, para lo cual contrata un estudio que determine un modelo que permita comprender de una forma más precisa el comportamiento de los ingresos del sector (I medidos en millones de dólares). Un investigador sugiere el siguiente modelo:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t \tag{2}$$

Donde, CE_t , CD_t , $LDies_t$, LEI_t y V_t representan el consumo de electricidad medido en millones de Kilovatios/hora, el consumo de diesel medido en millones de galones, el logaritmo del número de locomotoras diesel en el país, el número de locomotoras eléctricas y el número de viajeros (medido en miles de pasajeros) en el año t respectivamente.

Para efectuar este estudio se cuenta con la información para el período 1994-2003 (los datos se encuentran en el archivo T3-02-10.xls)

- a. De acuerdo con el enunciado anterior, estime el modelo 2 y reporte sus resultados en una tabla.
- b. Interprete los coeficientes estimados.

Pregunta 5

Continuando con la pregunta anterior, responda:

a. El modelo $I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t$ generó mucha discusión en la pequeña República y en diversos foros se sugirieron diferentes modelos:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \varepsilon_t \quad (3)$$

$$I_t = \alpha_1 + \alpha_2 LDies_t + \alpha_3 LEI_t + \varepsilon_t \quad (4)$$

$$I_t = \alpha_1 + \alpha_2 V_t + \varepsilon_t \quad (5)$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LEI_t + \varepsilon_t \quad (6)$$

$$I_t = \alpha_1 + \alpha_2 CD_t + \alpha_3 LDiest_t + \varepsilon_t \quad (7)$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LDiest_t + \alpha_4 LEI_t + \alpha_5 V_t + \varepsilon_t \quad (8)$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 V_t + \varepsilon_t \quad (9)$$

Determine cuál es el mejor modelo para explicar los ingresos del sector ferroviario en esta pequeña república (muestre todos los cálculos necesarios para tomar esta decisión y asegúrese que su respuesta está bien argumentada).

Pregunta 6

A partir del modelo que seleccionó en la pregunta anterior:

- Construya la tabla ANOVA.
- Analice la significancia individual de los coeficientes.
- Analice la significancia conjunta de los coeficientes.
- Determine cuál es el factor que más afecta el ingreso del sector ferroviario en esta nación.
- De acuerdo con su resultado, ¿qué puede sugerir al gobierno de esta nación para mejorar los ingresos del sector?

Taller 3: Regresión Múltiple

Econometría 06216

08-02-2010

Profesores: Julio César Alonso.

Monitoras: 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 AM del 09 de Agosto de 2010.
- Sólo se calificaran talleres en formato pdf. Cualquier otro formato no será tenido en cuenta.

Instrucciones:

- 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.
- Este taller debe ser escrito en computador.

Pregunta 1

Para las preguntas 1 a 3, realice los cálculos de manera manual. Presente todo el procedimiento y aclare muy bien cuáles fórmulas fueron empleadas y qué valores fueron reemplazados

Una reconocida distribuidora de elementos deportivos (patines, balones, bicicletas, accesorios, entre otros), supone que las ganancias de las ventas y_t (medidas en pesos) de sus productos sigue la siguiente relación:

$$y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \varepsilon_t \quad (1)$$

Donde X_{2t} corresponde a la cantidad de patines vendidos en el último mes; X_{3t} es el número de devoluciones mensuales por daño o imperfección; X_{4t} es el logaritmo del porcentaje de descuento ofrecido en el mes t . Además, ε_t representa una perturbación aleatoria que cumple los supuestos del teorema del Gauss-Markov.

Para 1000 observaciones se obtuvieron los siguientes valores:

$$\begin{array}{cccc}
 \sum_{t=1}^n X_{2t} = 4000 & \sum_{t=1}^n X_{3t} = 500 & \sum_{t=1}^n X_{4t} = 1917 & \\
 \sum_{t=1}^n (X_{2t})^2 = 10000 & \sum_{t=1}^n (X_{3t})^2 = 1000 & \sum_{t=1}^n (x_{4t})^2 = 4000 & \sum_{t=1}^n y_t = 5200 \\
 \sum_{t=1}^n X_{2t} \cdot X_{3t} = 2000 & \sum_{t=1}^n X_{2t} \cdot X_{4t} = 3200 & \sum_{t=1}^n X_{3t} \cdot X_{4t} = 800 & \sum_{t=1}^n y_t \cdot X_{3t} = 6400 \\
 \sum_{t=1}^n y_t \cdot X_{2t} = 11000 & \sum_{t=1}^n (y_t)^2 = 57098.36 & \sum_{t=1}^n y_t \cdot X_{4t} = 10500 &
 \end{array}$$

Construya las matrices $X^T X$, $X^T y$ y $y^T y$.

$$X^T X = \begin{pmatrix} 1000 & 4000 & 500 & 1917 \\ 4000 & 10000 & 2000 & 3200 \\ 500 & 2000 & 1000 & 800 \\ 1917 & 3200 & 800 & 4000 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 5200 \\ 11000 \\ 6400 \\ 10500 \end{pmatrix}$$

$$y^T y = 57098.36$$

Pregunta 2

Continuando con el enunciado de la pregunta anterior, estime el modelo 1.

$$\hat{\beta} = \begin{pmatrix} -1.586185 \\ -0.143025 \\ 5.5707899 \\ 2.3854413 \end{pmatrix}$$

Pregunta 3

Continuando con la pregunta anterior, responda:

a. Interprete los coeficientes estimados en la pregunta 3.

$\hat{\beta}_1 = -1.586185$: cuando el número de devoluciones por daño e imperfección y el número de patines vendidos en el último mes son cero y, cuando hay un descuento del 1%, la empresa sufre pérdidas por valor de 1.586185 pesos. (También era aceptable responder que no tiene interpretación económica.)

$\hat{\beta}_2 = -0.143025$: ante un aumento de una unidad en el número de patines vendidos en el último mes, las ganancias de la distribuidora disminuyen en 0.143025 pesos.

$\hat{\beta}_3 = 5.5707899$: ante un aumento de una unidad en el número de devoluciones por daño e imperfección en el último mes, las ganancias derivadas de las ventas de elementos deportivos aumenta en 5.57078 pesos.

$\hat{\beta}_4 = 2.3854413$: ante un aumento de 1% en el porcentaje de descuento ofrecido, las ganancias de las ventas de la distribuidora aumentan en 0.0238544 (2.38544/100) pesos.

b. Discuta la significancia conjunta de los coeficientes estimados

Para analizar la significancia conjunta de los coeficientes estimados es necesario utilizar una prueba de hipótesis de la forma $R\hat{\beta} = c$ donde $H_o : \hat{\beta}_2 = \hat{\beta}_3 = \hat{\beta}_4 = 0$ y $H_1 : \text{No } H_o$. Para probar esta hipótesis es necesario calcular un F, utilizando la fórmula: $F_c = \frac{(c-R\hat{\beta})^T (R(X^T X)^{-1} R^T)^{-1} (c-R\hat{\beta})/r}{SSE/n-k}$ donde $SSE = y^T y - \hat{\beta}^T X^T y$. Para este caso, se tiene que $r=4$ (número de iguales en la hipótesis nula) y $k=4$ (número de parámetros estimados).

Se encuentra que $F_c=1272.5018$. Como es un valor muy grande, se puede afirmar que se rechaza con un 99% de confianza. Así entonces, todos los coeficientes del modelo son estadísticamente diferentes de cero.

Pregunta 4

Para las preguntas 4 a 6, realice los cálculos de manera manual. Presente todo el procedimiento y aclare muy bien cuáles fórmulas fueron empleadas y que valores fueron reemplazados

El gobierno de una pequeña República está reconsiderando la viabilidad del transporte ferroviario, para lo cual contrata un estudio que determine un modelo que permita comprender de una forma más precisa el comportamiento de los ingresos del sector (I medidos en millones de dólares). Un investigador sugiere el siguiente modelo:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t \quad (2)$$

Donde, CE_t , CD_t , $LDies_t$, LEI_t y V_t representan el consumo de electricidad medido en millones de Kilovatios/hora, el consumo de diesel medido en millones de galones, el logaritmo del número de locomotoras diesel en el país, el número de locomotoras eléctricas y el número de viajeros (medido en miles de pasajeros) en el año t respectivamente.

Para efectuar este estudio se cuenta con la información para el período 1994-2003 (los datos se encuentran en el archivo T3-02-10.xls)

a. De acuerdo con el enunciado anterior, estime el modelo 1 y reporte sus resultados en una tabla. (Ver los talleres de semestres anteriores para un ejemplo del formato de tabla)

b. Interprete los coeficientes estimados.

La interpretación de los coeficientes estimados es la siguiente:

- $\hat{\alpha}_1 = 280106.177$. No tiene interpretación económica.
- $\hat{\alpha}_2 = 80.329$. Un aumento de un millón de Kilovatios/hora en el consumo de electricidad aumentará los ingresos en 80.329 millones de dólares.
- $\hat{\alpha}_3 = 49.363$. Un aumento de un millón de galones en el consumo de diesel aumentará los ingresos en 49.363 millones de dólares.
- $\hat{\alpha}_4 = -47.919$. Un aumento de 1% en el número de locomotoras diesel en el país, disminuirá los ingresos en 0.47919 (47.919/100) millones de dólares.
- $\hat{\alpha}_5 = -145.954$. Una locomotora eléctrica más en el país disminuirá los ingresos en 145.954 millones de dólares.
- $\hat{\alpha}_6 = 0.017$. Mil viajeros más aumentará los ingresos en 17.444 dólares.

Pregunta 5

Continuando con la pregunta anterior, responda:

a. El modelo $I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t$ generó mucha discusión en la pequeña República y en diversos foros se sugieron diferentes modelos:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \varepsilon_t \quad (3)$$

Table 1: Estimación del modelo (2)

VARIABLE DEPENDIENTE: It			
Estadísticos t entre paréntesis			
Ecuación 1			
1994-2003			
MCO			
constante	280106.177	(1.660)	*
CEt	80.329	(3.337)	***
CDt	49.363	(0.248)	
LDiestt	-47.919	(-0.914)	
LElt	-145.954	(-2.364)	**
Vt	0.017	(0.546)	
R2	0.9902		
R2 Ajustado	0.978		
F-global	81.17		***
de Obs.	10		

(*) nivel de significancia: 10%

(**) nivel de significancia: 5%

(***) nivel de significancia: 1%

MCO: Mínimos Cuadrados Ordinarios

Table 2: Estimación de los modelo (2) a (9)

		VARIABLE DEPENDIENTE: It							
		Estadísticos t entre paréntesis							
	Ecuación 1	Ecuación 2	Ecuación 3	Ecuación 4	Ecuación 5	Ecuación 6	Ecuación 7	Ecuación 8	
	1994-2003	1994-2003	1994-2003	1994-2003	1994-2003	1994-2003	1994-2003	1994-2003	1994-2003
	MCO	MCO	MCO	MCO	MCO	MCO	MCO	MCO	MCO
constante	280106.178 (1.660)	-93433.317 (-0.787)	926424.346 (8.388)	215324.455 (10.034)	224933.24 (1.762)	780596.889 (16.123)	255496.437 (2.076)	-63506.875 (-0.489)	
CEt	80.329 (3.337)	121.106 (7.119)	-	-	100.249 (6.963)	-	83.504 (4.544)	105.726 (3.828)	
CDt	49.363 (0.248)	-375.413 (-4.073)	-	-	-	327.112 (0.894)	-	-315.721 (-2.502)	
LDiest	-47.919 (-0.914)	-	-120.30 (-1.986)	-	-	-269.093 (-3.385)	-38.091 (-1.230)	-	
LElt	-145.954 (-2.364)	-	-175.021 (-1.443)	-	-207.962 (-6.129)	-	-140.369 (-2.709)	-	
Vt	0.017 (0.546)	-	-	0.201 (8.057)	-	-	0.0164 (0.574)	0.03 (0.723)	
R2	0.9902	0.9707	0.9214	0.8903	0.9845	0.9084	0.990	0.973	
R2 Ajustado	0.978	0.9623	0.8989	0.8766	0.9801	0.8823	0.9822	0.9596	
F-global	81.17	115.92	41.01	64.92	222.11	34.72	124.880	72.19	
de Obs.	10	10	10	10	10	10	10	10	

(*) nivel de significancia: 10%
 (**) nivel de significancia: 5%
 (***) nivel de significancia: 1%
 MCO: Mínimos Cuadrados Ordinarios

$$I_t = \alpha_1 + \alpha_2 LDiest_t + \alpha_3 LEI_t + \varepsilon_t \tag{4}$$

$$I_t = \alpha_1 + \alpha_2 V_t + \varepsilon_t \tag{5}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LEI_t + \varepsilon_t \tag{6}$$

$$I_t = \alpha_1 + \alpha_2 CD_t + \alpha_3 LDiest_t + \varepsilon_t \tag{7}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LDiest_t + \alpha_4 LEI_t + \alpha_5 V_t + \varepsilon_t \tag{8}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 V_t + \varepsilon_t \tag{9}$$

Determine cuál es el mejor modelo para explicar los ingresos del sector ferroviario en esta pequeña república (muestre todos los cálculos necesarios para tomar esta decisión y asegúrese que su respuesta está bien argumentada).

Para determinar cuál es el mejor modelo de los ocho a considerar, podemos emplear el R^2 ajustado de cada modelo. De acuerdo con estos resultados, el mejor modelo corresponde al número 8, pues es el que posee el R^2 ajustado mayor entre todos los modelos.

Pregunta 6

A partir del modelo que seleccionó en la pregunta anterior:

a. Construya la tabla ANOVA.

La tabla ANOVA del modelo seleccionado es la siguiente:

Table 3: Tabla ANOVA del modelo seleccionado

Fuente de la Variación	SS	Grados de libertad	MS	F-global
Regresión	22596488607.1	4	5649122152	124.8806057
Error	226180923.8	5	45236184.76	
Total	22822669530.9	9		

Table 4: Cálculo de los coeficientes estandarizados

Variable	MCO	Coefficiente Estandarizado	Desviación estándar
It	NA	NA	6725.785067
CEt	83.504	0.228152609	18.37643
LDiest	-38.0911	-0.175358656	30.96326
LEIt	-140.3697	-1.081297803	51.81016
Vt	0.01639	6.95732E-08	0.02855

NA: No Aplica

MCO: Mínimos cuadrados ordinarios

b. Analice la significancia individual de los coeficientes.

Para probar individualmente la hipótesis nula de que $\alpha_i = 0$ versus la hipótesis alterna $\alpha_i \neq 0$ para $i = 1, 2, \dots, 5$ podemos remitirnos a los estadísticos t. Como se puede observar, individualmente $\hat{\alpha}_3$ y $\hat{\alpha}_5$ no son significativos. Por otra parte, los coeficientes asociados al consumo de energía y el número de locomotoras eléctricas son significativos a un nivel de significancia del 1%.

c. Analice la significancia conjunta de los coeficientes.

Para comprobar la hipótesis nula de que $\alpha_2 = \alpha_3 = \dots = \alpha_5 = 0$ versus la hipótesis alterna no H_0 , se puede emplear el estadístico F-global que en este caso corresponde a 124.8806 el cual es mayor al valor crítico de la distribución F (11.39) con 4 y 5 grados de libertad en el numerador y denominador, respectivamente. Por tanto se puede rechazar la hipótesis nula de que todos los coeficientes asociados a pendientes son conjuntamente iguales a cero. Luego los coeficientes son conjuntamente significativos.

d. Determine cuál es la variable que más afecta el ingreso del sector ferroviario en esta nación.

Para poder efectuar esta comparación, debemos emplear los coeficientes estandarizados que se calculan de la siguiente forma: $\hat{\alpha}_i^E = \hat{\alpha}_i \frac{s_{X_i}}{s_y}$ para $i = 2, 3, \dots, 5$. En el cuadro se presentan los resultados de estandarizar los coeficientes. Noten que el coeficiente estandarizado más grande es el asociado con el consumo de energía. Así esta es la variable que afecta más el ingreso ferroviario.

e. De acuerdo con su resultado, ¿qué puede sugerir al gobierno de esta nación para mejorar los ingresos del sector?

Noten que parece paradójico que a mayor consumo de energía mayores ingresos, pero esto puede estar asociado a que las locomotoras eléctricas son las que generan más ingresos para el sector. Además, parece razonable concluir, dado el bajo coeficiente estandarizado asociado a los viajeros, que los ingresos son generados por transporte de mercancías. Así, será importante para el gobierno tener en cuenta los efectos del transporte de mercancías para este sector. Sería conveniente replantear el modelo, empleando información de carga transportada por este sector para poder llegar a recomendaciones más concretas.

Año	CE (T)	CD (T)	I (T)	Ldies (T)	LEI (T)	V (T)
1994	4734	480	306216	2325	1869	581940
1995	4842	444	322113	2223	1863	570852
1996	5058	450	340563	2199	1863	582795
1997	5010	441	349986	2073	1818	545673
1998	5262	432	389574	2082	1779	823050
1999	5454	426	398838	1914	1776	948981
2000	5502	399	426720	1866	1674	1075845
2001	5214	360	407982	1740	1704	1018248
2002	5379	333	418398	1740	1644	1054590
2003	5526	333	462069	1623	1503	1096509