Taller #5
Regresión Simple
Econometría 06169

Profesores: Julio César Alonso
Diego Yépez

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de agosto.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

1. Demuestre que: $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$.
2. Demuestre que: $\sum_{i=1}^{n} \hat{\varepsilon}_i x_i = 0$.

En una encuesta realizada a 40 familias se ha obtenido la información del gasto anual en productos lácteos (G_i) y la renta disponible (Y_i) que se reportan en el archivo T5-02-04.xls. Para evitar distorsiones producidas por los diferentes tamaños de los hogares, tanto el consumo como la renta se han expresado en términos *per capita* (miles de pesos por persona). Un investigador no esta seguro cual de los dos modelos siguientes es más adecuado para determinar la curva de Engel de los productos agrícolas:

$$\ln(G_i) = \alpha + \beta Y_i + \varepsilon_i \qquad (1)$$

$$\ln(G_i) = \delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i \qquad (2)$$

3. Empleando la información anterior:
   a. Estime (1) y (2) y reporte sus resultados en una tabla.
   b. ¿Cuál de los dos modelos es mejor? Explique su decisión
4. Empleando el modelo que seleccionó en la pregunta anterior, responda:
   a. Interprete los coeficientes estimados.
   b. Esboce una gráfica de la curva de Engel para los productos lácteos.
5. El ingreso medio *per capita* de la ciudad donde fue realizada la encuesta es de \$1.500.000 por persona, responda:
   a. ¿Cuál es el gasto *per capita* medio esperado? Reporte también un intervalo de confianza con el 99% de confianza para este gasto *per capita* medio esperado.
   b. Si un vendedor llegase a una casa con ingreso anual de 15 mil pesos *per capita*, ¿cuánto debería esperar vender en productos lácteos? Construya un intervalo de confianza con el 99% de confianza para su respuesta.

Taller #5
Respuestas Sugeridas
Regresión Simple
Econometría 06169

Julio César Alonso

Notas:

- Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de agosto.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

1. Demuestre que: $\sum_{i=1}^{n}\hat{\varepsilon}_i = 0$.

Recuerden que $\hat{\varepsilon}_i = y_i - \hat{y}_i$ y $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$, por tanto tenemos que:

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}\left(y_i - \left(\hat{\beta}_1 + \hat{\beta}_2 x_i\right)\right)$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\beta}_1 - \sum_{i=1}^{n}\hat{\beta}_2 x_i$$

Empleando **¡Error! No se encuentra el origen de la referencia.** y **¡Error! No se encuentra el origen de la referencia.**,

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\left(\overline{y} - \hat{\beta}_2\overline{x}\right) - \hat{\beta}_2\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\overline{y} + \sum_{i=1}^{n}\hat{\beta}_2\overline{x} - \hat{\beta}_2\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - n\overline{y} + n\hat{\beta}_2\overline{x} - \hat{\beta}_2\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - n\frac{\sum_{i=1}^{n} y_i}{n} + n\hat{\beta}_2\frac{\sum_{i=1}^{n} x_i}{n} - \hat{\beta}_2\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i + \hat{\beta}_2\sum_{i=1}^{n} x_i - \hat{\beta}_2\sum_{i=1}^{n} x_i = 0$$

Q.E.D.

Noten que este resultado no depende de ningún supuesto. Además este resultado implica que:

$$\overline{\hat{\varepsilon}}_i = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i}{n} = 0$$

2. Demuestre que: $\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = 0$.

Prueba:

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n}(y_i - \hat{y}_i)x_i = \sum_{i=1}^{n}(y_i x_i - \hat{y}_i x_i)$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\hat{y}_i x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\left(\hat{\beta}_1 + \hat{\beta}_2 x_i\right)x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\hat{\beta}_1 x_i - \sum_{i=1}^{n}\left(\hat{\beta}_2 x_i\right)x_i$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\hat{\beta}_1 x_i - \sum_{i=1}^{n}\hat{\beta}_2 x_i^2$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i - \hat{\beta}_2\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \left(\overline{y} - \hat{\beta}_2\overline{x}\right)\sum_{i=1}^{n} x_i - \hat{\beta}_2\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \overline{y}\sum_{i=1}^{n} x_i + \hat{\beta}_2\overline{x}\sum_{i=1}^{n} x_i - \hat{\beta}_2\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - \overline{y}n\overline{x} + \hat{\beta}_2\overline{x}n\overline{x} - \hat{\beta}_2\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x} + \hat{\beta}_2\left(n\overline{x}^2 - \sum_{i=1}^{n} x_i^2\right)$$

Remplazando obtenemos:

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x} + \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\left(n\overline{x}^2 - \sum_{i=1}^{n} x_i^2\right)$$

$$\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = \sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x} - \left(\sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x}\right) = 0$$

Q.E.D.

Noten que este resultado implica que:

$$Cov\left(\hat{\varepsilon}_i, x_i\right) = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i x_i - \overline{x}\cdot\overline{\hat{\varepsilon}}_i}{n-1} = 0$$

Este resultado tampoco depende de ningún supuesto y por tanto siempre se cumplirá.

En una encuesta realizada a 40 familias se ha obtenido la información del gasto anual en productos lácteos $\left(G_i\right)$ y la renta disponible $\left(Y_i\right)$ que se reportan en el archivo T5-02-04.xls. Para evitar distorsiones producidas por los diferentes tamaños de los hogares, tanto el consumo como la renta se han expresados en términos *per capita* (miles de pesos por persona). Un investigador no esta seguro cual de los dos modelos siguientes es más adecuado para determinar la curva de Engel de los productos agrícolas:

$$\ln\left(G_i\right) = \alpha + \beta Y_i + \varepsilon_i \qquad (1)$$

$$\ln\left(G_i\right) = \delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i \qquad (2)$$

3. Empleando la información anterior:
   a. Estime (1) y (2) y reporte sus resultados en una tabla.

Los resultados se reportan en la **Tabla 1**.

Tabla 1. Resultados de la estimación de los modelos (1) y (2).

	VARIABLE DEPENDIENTE: Ln(G_i) Estadísticos t entre paréntesis	
	Ecuación 1 MCO	Ecuación 2 MCO
Constante	1,69 (14,29) ***	3,049 (26,06) ***
Y_i	0,00048 (6,14) **	---
1/(Y_i)	---	-822,024 (-6,21) **
S^2	0,0737	0,0728
R^2	0,4978	0,5040
# de Obs.	40	40

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

b. ¿Cuál de los dos modelos es mejor? Explique su decisión

Noten que para ambos modelos el intercepto y respectiva pendientes son individualmente significativos. Además si observamos detalladamente ambas regresiones son más o menos similares en cuanto a su ajuste. Eso lo podemos ver en el coeficiente de determinación tan cercano en ambos casos. Es más si graficamos los puntos y respectiva línea de regresión en ambos casos, no parece ser un modelo mejor que el otro.

Pero como necesitamos tomar una decisión, podemos emplear el R^2 para determinar cuál de los dos modelos explica mejor la variabilidad total del logaritmo del gasto en lácteos. Así, el mejor modelo que explica los datos es el (2).

Gráfico 1 Logaritmo del gasto en lácteos versus el inverso del ingreso

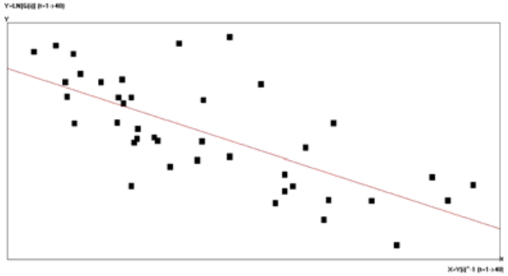
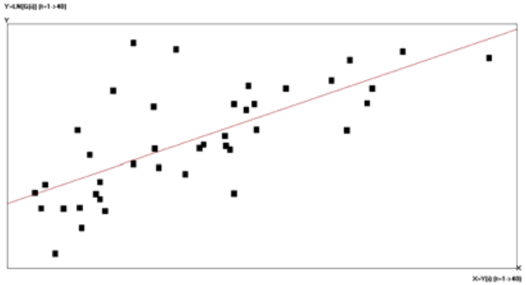


Gráfico 2 Logaritmo del gasto en lácteos versus el ingreso



4. Empleando el modelo que seleccionó en la pregunta anterior, responda:
   a. Interprete los coeficientes estimados.

Dado que el modelo seleccionado corresponde a: $\ln(G_i) = \delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i$. Noten que (2) equivale a:

$$G_i = e^{\delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i} \qquad (3)$$

Para interpretar el significado de los coeficientes estimados, podemos derivar **¡Error! No se encuentra el origen de la referencia.** con respecto a la variable independiente. Es decir:

$$\frac{\partial G_i}{\partial Y_i} = -\gamma \frac{1}{Y_i^2} e^{\delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i} = -\gamma \frac{1}{Y_i^2} G_i \qquad (4)$$

$$\gamma = -\frac{\partial G_i / G_i}{\partial Y_i / Y_i^2} = -\frac{\partial G_i / G_i}{\partial Y_i / Y_i^2}\frac{100}{100} = -Y_i \frac{\Delta\% G_i}{\Delta\% Y_i} \qquad (5)$$

$$\frac{\Delta\% G_i}{\Delta\% Y_i} = -\frac{\gamma}{Y_i} \qquad (6)$$

Así, un aumento del 1% en el ingreso provocará un cambio de $-\gamma_3/Y_i$ % en el gasto anual en lácteos. Es decir el efecto de aumentar el ingreso no es constante y dependerá del nivel de ingresos.

Así, $\hat{\gamma} = $-822.024 corresponderá a parte de la elasticidad ingreso del gasto en lácteos. Es más esta elasticidad depende del nivel de ingresos de cada familia, de tal forma que a mayor nivel de ingresos, menor será (el valor absoluto) de la elasticidad.

Por otro lado, $\hat{\delta}$ no tiene interpretación económica.

b. Esboce una gráfica de la curva de Engel para los productos lácteos.

Antes de graficar la curva de Engel recuerden que la curva de Engel es una curva que relaciona las cantidades demandas de un bien y el nivel de ingresos. En este caso estamos estudiando el gasto en productos lácteos, y no exactamente las cantidades demandadas, así que en este caso graficaremos la curva de Engel para el gasto en lácteos. Primero, note que a partir de (3) se pueden determinar las características de la curva de Engel. Por ejemplo. cuando el ingreso es lo suficientemente grande, entonces las cantidades demandas tienden a:

$$\lim_{Y_i \to \infty} G_i = \lim_{Y_i \to \infty}\left[e^{\delta + \gamma\left(\frac{1}{Y_i}\right) + \mu_i}\right] = e^{\delta + \mu_i} \qquad (7)$$

Segundo, debemos determinar la pendiente de la curva y su curvatura. De nuestros cálculos anteriores sabemos que el signo de la pendiente $\frac{\partial G_i}{\partial Y_i} = -\gamma \frac{1}{Y_i^2} G_i$ dependerá del signo de γ y la curvatura de la curva de Engel dependerá de la segunda derivada ($\frac{\partial^2 G_i}{\partial Y_i^2} = 2\gamma \frac{G_i}{Y_i^3} + \gamma^2 \frac{G_i}{Y_i^4}$). Así, si γ es menor que cero, entonces la segunda derivada podrá ser positiva o negativa. En especial, ésta será negativa si:

$$\frac{\partial^2 G_i}{\partial Y_i^2} = 2\gamma \frac{G_i}{Y_i^3} + \gamma^2 \frac{G_i}{Y_i^4} < 0$$

$$\gamma \frac{G_i}{Y_i^3}\left[2 + \gamma \frac{1}{Y_i}\right] < 0$$

Dado que $\gamma < 0$, entonces tendremos que

$$\left[2+\gamma\frac{1}{Y_i}\right]>0$$

Es decir,

$$\gamma < -2Y_i \qquad (8)$$

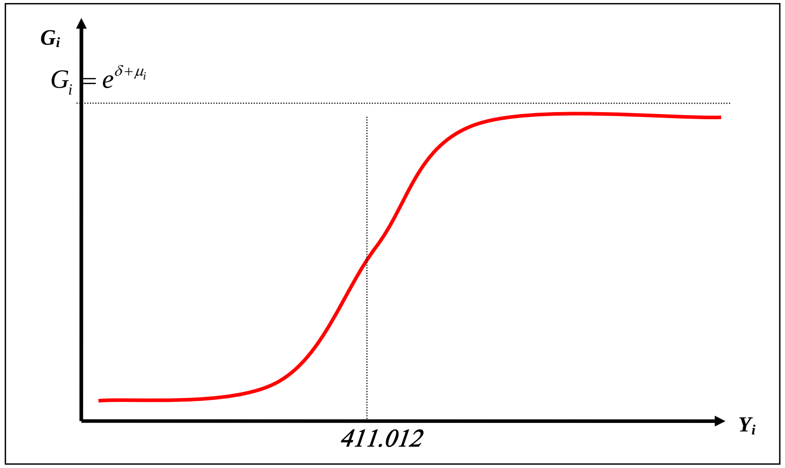
y por tanto:

$$Y_i < -\frac{\gamma}{2} \qquad (9)$$

En este caso, tendremos que para $Y_i < -\frac{-822.024}{2} = 411.012$ la segunda derivada será negativa y para valores menores a este ingreso será positiva.

En conclusión, dado que $\hat{\gamma} = -822.024$ es negativo, entonces la pendiente de la curva de Engel será negativa y la segunda derivada será negativa para $Y_i < 411.012$ y positive en caso contrario. Noten que $\gamma_3 < 0$ implicará que un aumento del ingreso aumentará el gasto pero cada vez este aumentara menos a medida que el ingreso aumenta. En otras palabras se tratará de un bien normal. Es decir la curva de Engel tendrá que verse como en la Figura 1.

Figura 1. Curva de Engel con $\gamma_3 < 0$ **.**



5. El ingreso medio *per capita* de la ciudad donde fue realizada la encuesta es de $1.500.000 por persona, responda:

a. ¿Cuál es el gasto *per capita* medio esperado? Reporte también un intervalo de confianza con el 99% de confianza para este gasto *per capita* medio esperado.

El gasto *per capita* medio esperado corresponde a:

$$\ln\left(\hat{G}_i\right) = \hat{\delta} + \hat{\gamma}\left(\frac{1}{Y_i}\right) = 3.049 - 822.024\left(\frac{1}{Y_i}\right)$$

Así, para un ingreso per cápita de $1.500.000 por persona tendremos:

$$\ln\left(\hat{G}_i\right) = 3.049 - 822.024\left(\frac{1}{1500}\right) = 2.50$$

por lo tanto,

$$\hat{G}_i = e^{2.5} = 12.19$$

Noten que para proveer un intervalo de confianza par este valor con el 99% de confianza debemos hacerlo a partir de su logaritmo. Es decir:

$$\hat{G}_i \pm t_{\alpha/2,n-2}\, s\sqrt{\frac{1}{n}+\frac{\left(x_p-\overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}} = 2.5 \pm (2.71)(0.27)\sqrt{\frac{1}{40}+\frac{\left(\frac{1}{1500}-0.0008\right)^2}{0.000004}}$$

$$2.5 \pm 0.13 = [2.37, 2.63]$$

Ahora para conocer un intervalo de confianza para el gasto, tenemos que:

$$\left[e^{2.37}, e^{2.63}\right] = [10.72, 13.87]$$

b. Si un vendedor llegase a una casa con ingreso anual de $1.500.000 *per capita*, ¿cuánto debería esperar vender en productos lácteos? Construya un intervalo de confianza con el 99% de confianza para su respuesta.

En este caso, el valor esperado de la realización es igual que antes, es decir: $\hat{G}_i = e^{2.5} = 12.19$.

El intervalo de confianza para el logaritmo del gasto está dado por:

$$\hat{G}_i \pm t_{\alpha/2,n-2}\, s\sqrt{1+\frac{1}{n}+\frac{\left(x_p-\overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}} = 2.5 \pm (2.71)(0.27)\sqrt{1+\frac{1}{40}+\frac{\left(\frac{1}{1500}-0.0008\right)^2}{0.000004}}$$

$$2.5 \pm 0.74 = [1.76, 3.24]$$

Ahora para conocer un intervalo de confianza para el gasto, tenemos que:

$$\left[e^{1.76}, e^{3.24}\right] = [5.8, 25.63]$$