

Tema 5

Modelos no lineales y transformación de variables

(actualizadas el 07-07-2023)

Tema 5. Modelos no lineales y transformación de variables

5.1 Unidades de medida

5.2 Formas funcionales

5.3 Selección de modelos: AIC

Bibliografía

- Ezequiel Uriel (2013): Capítulo 2 (epígrafe 2.4) y Capítulo 3 (epígrafes 3.4 y 3.5)
- Wooldridge (2015): Capítulo 6 (epígrafes 6.1 y 6.2)
- Stock y Watson (2012): Capítulo 8 (epígrafe 8.2)

5.1 Unidades de medida

¿Qué ocurre si en lugar de medir el salario en euros pasamos a medirlo en miles de euros?

Unidades de medida: cambios de escala

1. Cambios de escala en los **regresores** (x)

- Si la variable exógena x cambia de escala, **solo cambia su coeficiente**
- En concreto, si x se **multiplica** por una constante ($c \neq 0$), su coeficiente se **divide** por la misma constante
- La estimación de la ordenada en el origen no cambia.

2. Cambios de escala en el **regresando** (y)

- Si la variable endógena y cambia de escala, **todos los coeficientes cambian**
- En concreto, si y se **multiplica** por una constante todos los coeficientes de las x se **multiplican** por la misma constante
- La ordenada también se ve afectada por el cambio de escala

3. Los indicadores de bondad de ajuste y los contrastes de significatividad no cambian

Cambios de escala (ejemplo)

Modelo 1: $Precio(€) = \beta_1 + \beta_2 Superficie(m^2) + \beta_3 Habitaciones + u$

Modelo 2: $Precio(1000€) = \beta_1 + \beta_2 Superficie(m^2) + \beta_3 Habitaciones + u$

Modelo 3: $Precio(€) = \beta_1 + \beta_2 Superficie(pies^2) + \beta_3 Habitaciones + u$

Variables	Modelo 1	Modelo 2	Modelo 3
Contante	-18.934,9 (34.780,6)	-18,9349 (34,7806)	-18.934,9 (34.780,6)
Superficie (m ²)	3.674,9 (368,8)	3,6749 (0,3688)	
Superficie (pies ²)			341,4 (34,3)
Habitaciones	-7.688,8 (8.435,2)	-7,6888 (8,4352)	-7.688,8 (8.435,2)
R ²	0.398080	0.398080	0.398080
Valor p est. F	6,86·10 ⁻³⁶	6,86·10 ⁻³⁶	6,86·10 ⁻³⁶

1 pie cuadrado =
0,092903 metros
cuadrados

Unidades de medida: cambios de origen

- **NO** afectan a los coeficientes de pendiente
- **SÍ** afectan a β_1
- Los indicadores de bondad de ajuste y los contrastes de significatividad se mantienen.

Modelo 1: $Precio = \beta_1 + \beta_2 Superficie + u$

Modelo 2: $Precio = \beta_1 + \beta_2 (Superficie - \overline{Superficie}) + u$

Modelo 3: $Precio - \overline{Precio} = \beta_1 + \beta_2 (Superficie - \overline{Superficie}) + u$

Variables	Modelo 1	Modelo 2	Modelo 3
Contante	-42.797,3 (22.892,6)	272593 (7.138,8)	0,0 (7.138,8)
Superficie	3.416,4 (235,6)	3.416,4 (235,6)	3.416,4 (235,6)
R ²	0.396513	0.396513	0.396513
Valor p est. F	$5,69 \cdot 10^{-37}$	$5,69 \cdot 10^{-37}$	$5,69 \cdot 10^{-37}$

5.2 Formas funcionales

Muchas veces los datos se transforman tomando logaritmos ...

El modelo lineal, ¿no es muy restrictivo?

- En Economía hay muchas relaciones que pueden ser no lineales ¿Podemos tratarlas/modelizarlas con nuestro MRLS?
- La respuesta es **SÍ**, al menos en muchos casos, simplemente tendremos que redefinir las variables dependiente e independiente
- Por ejemplo:

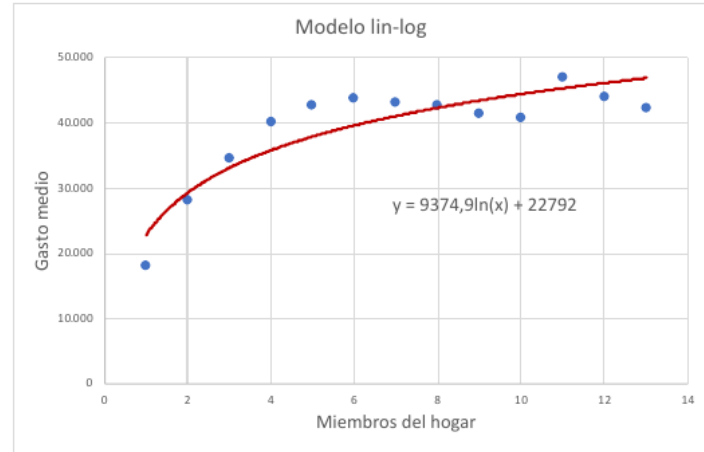
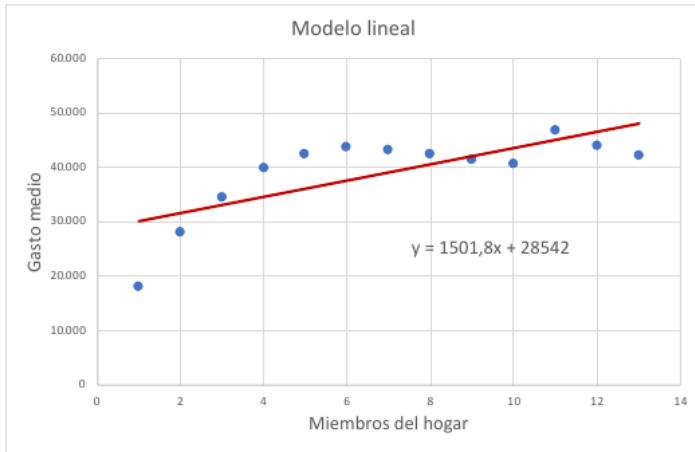
$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + u$$

$$\log(y) = \beta_1 + \beta_2 x + u$$

- ¿Cómo se interpretan los parámetros de los anteriores modelos? Lo vemos enseguida

a veces los datos no siguen una relación lineal

- A veces suponer que la relación entre x e y es una recta no es la mejor manera de hacer el ajuste a una nube de puntos no es lo más adecuado.



- "Incluso" a veces no tenemos clara cómo es la relación entre x e y
- Un ejemplo de esto último [aquí](#)

Las 4 formas funcionales más habituales

- **Modelo lineal-lineal:** $y = \beta_1 + \beta_2 x + u$ β_2 es el **efecto marginal**
- **Modelo log-log:** $\log(y) = \beta_1 + \beta_2 \log(x) + u$ β_2 es la **elasticidad**
- **Modelo log-lin:** $\log(y) = \beta_1 + \beta_2 x + u$ β_2 es una semi-elasticidad
- **Modelos lin-log:** $y = \beta_1 + \beta_2 \log(x) + u$ β_2 es una semi-elasticidad

modelo lineal-lineal

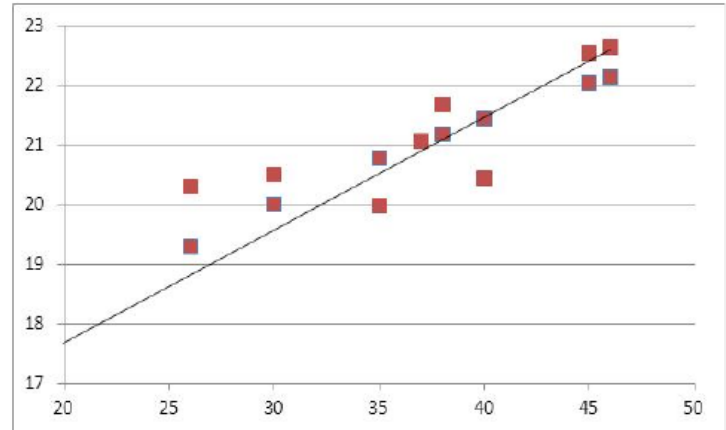
$$y = \beta_1 + \beta_2 x + u$$

Ajuste lineal: $y_i = \beta_1 + \beta_2 x_i + u_i$

La variación de y generada por una variación en x , es decir, la **propensión marginal**, es:

$$\frac{\delta y}{\delta x} = \beta_2$$

La pendiente de la recta nos da una medida de las unidades de cambio en y_i cuando la x_i se incrementa en una unidad.



Si x cambia 1 unidad ($\Delta x = 1$) \longrightarrow y cambiará β_2 unidades ($\Delta y = \beta_2$)

modelo doblemente logarítmico

$$\log(y) = \beta_1 + \beta_2 \log(x) + u$$

Ajuste logarítmico: $\ln(y_i) = \beta_1 + \beta_2 \ln(x_i) + u_i$

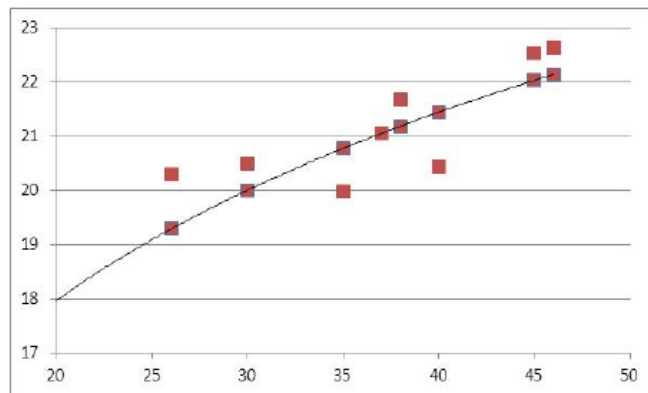
Ahora

$$\frac{\delta \ln(y)}{\delta \ln(x)} = \beta_2$$

*La interpretación del coeficiente se relaciona con el concepto de **elasticidad**:*

$$e_{xy} = \frac{\partial y / y}{\partial x / x} = \frac{\partial y}{\partial x} \frac{x}{y}$$

Si x cambia 1% ($\Delta x / x = 1\%$) \longrightarrow y cambiará $\beta_2\%$ ($\Delta y / y = \beta_2\%$)



modelos log-lin y lin-log

Ajuste logarítmico-lineal: $\ln(y_i) = \beta_1 + \beta_2 x_i + u_i$

Si x cambia 1 unidad  y cambiará $100 \cdot \beta_2 \%$

Ajuste lineal-logarítmico: $y_i = \beta_1 + \beta_2 \ln(x_i) + u_i$

Si x cambia 1%  y cambiará $\beta_2/100$ unidades

Formas funcionales (resumen)

CUADRO 2.5. Interpretación de en los diferentes modelos.

Modelo	Si x aumenta en	entonces y
		se incrementará en
lineal	1 unidad	$\hat{\beta}_2$ unidades
lineal logarítmico	1%	$(\hat{\beta}_2 / 100)$ unidades
logarítmico lineal	1 unidad	$(100\hat{\beta}_2)\%$
doblemente logarítmico	1%	$\hat{\beta}_2\%$

Formas funcionales (ejemplo)

- Lineal: $\widehat{Precio} = -42797 + 3416Superficie$
 - En promedio, si la superficie se incrementa en 1 metro cuadrado, el precio lo hace en 3.416€. **Propensión marginal**
- Doblemente logarítmico: $\ln(\widehat{Precio}) = 8,5 + 0,87\ln(Superficie)$
 - En promedio, si la superficie se incrementa en 1%, el precio lo hace en 0,87%. **Elasticidad**
- Lineal logarítmico: $\widehat{Precio} = -946506 + 272701\ln(Superficie)$
 - En promedio, si la superficie se incrementa en 1%, el precio lo hace en 2.727,01€ (272701/100).
- Logarítmico lineal: $\ln(\widehat{Precio}) = 11,4 + 0,01Superficie$
 - En promedio, si la superficie se incrementa en 1 metro cuadrado, el precio lo hace en 1% (0.01·100%).

5.3 Selección de modelos: AIC

AIC es un estadístico que se suele utilizar para comparar modelos

Si estimamos varios modelos ¿con cuál nos quedamos?

- Puede depender de muchas cosas, pero al menos hay que conocer el estadístico AIC

$$AIC = -\frac{2}{N} \ln L(y, \tilde{\beta}) + \frac{2k}{N}$$

$$AIC = 1 + \ln(2\pi) + \ln\left(\frac{SCR}{N}\right) + \frac{2(k+1)}{N}$$

Algunas ideas sobre el AIC

- **CRITERIO:** Menor valor del AIC implica un mejor “ajuste” del modelo.
- No es una medida de carácter relativo como el R^2 . Su valor no indica por sí mismo “un ajuste elevado o reducido”, pero sí permite comparar modelos.
- AIC puede aplicarse a modelos sin término constante.
- AIC penaliza la introducción de regresores adicionales; por lo tanto, permite comparar modelos con distinto número de regresores
- El AIC tampoco se puede utilizar para comparar modelos con distinto regresando. Sin embargo es posible encontrar una transformación del mismo que permita la comparación.
 - Por ejemplo si tenemos el modelo 1 (con regresando y) y el modelo 2 (regresando $\log(y)$) se pueden comparar el AIC del modelo 1 con el AIC transformado del modelo 2: $AIC'_2 = AIC_2 + 2 N \log(y)$

Comparación de modelos con AIC (ejemplo)

Modelos alternativos:

$$1) \text{ Precio} = \beta_1 + \beta_2 \text{Superficie} + u$$

$$2) \text{ Precio} = \beta_1 + \beta_2 \ln(\text{Superficie}) + u$$

$$3) \text{ Precio} = \beta_1 + \beta_2 \text{Superficie} + \beta_3 \text{Habitaciones} + u$$

$$4) \text{ Precio} = \beta_2 \text{Superficie} + \beta_3 \text{Habitaciones} + u$$

$$5) \text{ Precio} = \beta_1 + \beta_2 \text{Superficie} + \beta_3 \text{Edad} + u$$

$$6) \ln(\text{Precio}) = \beta_1 + \beta_2 \text{Superficie} + u$$

$$7) \ln(\text{Precio}) = \beta_1 + \beta_2 \text{Superficie} + \beta_3 \text{Habitaciones} + u$$

$$8) \ln(\text{Precio}) = \beta_2 \text{Superficie} + \beta_3 \text{Habitaciones} + u$$

$$N = 322, \quad \overline{\ln(\text{Precio})} = 12,376$$

Transparencia de Ezequiel Uriel

Medidas de Bondad de Ajuste de 8 modelos

Modelos	1	2	3	4	5	6	7	8
Regresando	Precio	Precio	Precio	Precio	Precio	Ln(Precio)	Ln(Precio)	Ln(Precio)
Regresores	Intercepto Superficie	Intercepto Ln(Superficie)	Intercepto Superficie Habitaciones	Superficie Habitaciones	Intercepto Superficie Edad	Intercepto Superficie	Intercepto Superficie Habitaciones	Superficie Habitaciones
R ²	0,3965	0,3148	0,3981	0,8393	0,4223	0,3856	0,3872	0,9627
R ² ajustado	0,3946	0,3127	0,3943	0,8388	0,4188	0,3837	0,3833	0,9626
AIC	8489,6	8530,5	8490,8	8489,1	8477,0	324,51	325,70	1479,5
AIC corregido						8294,7	8295,8	9449,6