

Sesión 26 (C) — Leer una tabla de resultados: cuatro modelos, cuatro lecturas

Marcos Bujosa

Descripción de la práctica

Esta práctica tiene poco código y mucho texto. Estima cuatro modelos con datos que el curso ya ha usado y, para cada tabla de resultados, pide un párrafo que diga lo que cada coeficiente dice y lo que no dice, con las reglas que el curso ha ido fijando: las unidades de cada variable (lección 12), la lectura de los logaritmos en tanto por uno y su versión exacta (lecciones 15 y 16), “con lo demás fijo” (lección 10), la constante solo si el origen de los regresores no constantes tiene datos cerca (lecciones 7 y 16), el lenguaje de los contrastes sin afirmaciones tajantes sobre los parámetros (lecciones 12 y 13) y la diferencia entre lo que el modelo mide y un efecto causal (lección 9). Leer una tabla es el ejercicio que cierra el curso: el que hay que saber hacer con cualquier tabla de resultados.

Para cada tabla, escriba el párrafo antes de leer “Lo que debe observar” y compárelo después con el párrafo modelo de la sección “Respuestas”.

Objetivos

1. Leer un modelo en niveles con regresores en unidades muy distintas.
2. Leer un modelo log-nivel con una indicadora.
3. Leer un modelo nivel-log con una pregunta de política detrás.
4. Leer un modelo log-log con elasticidades y distinguir asociación de causalidad.

Comandos nuevos de esta práctica. No hay ninguno: `open` (sesión 4), `logs` ([actividad 1 \(html\)](#) de la práctica B de la sesión 25), `summary --simple` y `printf` (sesión 7), `ols` y sus accesorios (sesiones 11 a 18), la función `corr`, como el comando `corr` de la sesión 7, y la función `critical` (sesión 18).

Tabla 1 - El precio de una vivienda en niveles (`hprice1`)

Abra `hprice1` (pestaña Wooldridge; es el fichero de la [actividad 1 \(html\)](#) de la práctica D de la sesión 22) y estime el precio sobre la superficie construida, el número de dormitorios y la superficie de la parcela.

`summary --simple` da las unidades y los rangos: el precio está en miles de dólares, las dos superficies en pies cuadrados. El `ols` estima el modelo (Modelo $\rightarrow$ Mínimos cuadrados ordinarios) y el `printf` traduce dos coeficientes a unidades más cómodas: cien pies cuadrados de casa y mil pies cuadrados de parcela.

o bien teclee en línea de comandos:

```
open hprice1 --quiet
```

```
summary price sqrft bdrms lotsize --simple
ols price const sqrft bdrms lotsize
printf "100 pies cuadrados mas de casa: %.1f mil dolares ; 1000 pies cuadrados mas de parcela: %.2f mil dolares\n",
↪ 100*$coeff(sqrft), 1000*$coeff(lotsize)
```

Licencia: Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

	Media	Mediana	D. T.	Min	Máx
price	293,5	265,5	102,7	111,0	725,0
sqrft	2014	1845	577,2	1171	3880
bdrms	3,568	3,000	0,8414	2,000	7,000
lotsize	9020	6430	10174	1000	92681

Modelo 1: MCO, usando las observaciones 1-88
Variable dependiente: price

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-21,7703	29,4750	-0,7386	0,4622	
sqrft	0,122778	0,0132374	9,275	1,66e-14	***
bdrms	13,8525	9,01015	1,537	0,1279	
lotsize	0,00206771	0,000642126	3,220	0,0018	***
Media de la vble. dep.	293,5460	D.T. de la vble. dep.	102,7134		
Suma de cuad. residuos	300723,8	D.T. de la regresión	59,83348		
R-cuadrado	0,672362	R-cuadrado corregido	0,660661		
F(3, 84)	57,46023	Valor p (de F)	2,70e-20		
Log-verosimilitud	-482,8775	Criterio de Akaike	973,7549		
Criterio de Schwarz	983,6643	Crit. de Hannan-Quinn	977,7472		

Sin considerar la constante, el valor p más alto fue el de la variable 3 (bdrms)

100 pies cuadrados mas de casa: 12,3 mil dolares ; 1000 pies cuadrados mas de parcela: 2,07 mil dolares

Escriba: un párrafo con la lectura de los cuatro coeficientes, en el que conste de qué unidades se habla, qué se mantiene fijo, qué dice el t de **bdrms** y si la constante se lee.

Lo que debe observar

Tres trampas. La primera son las unidades: el coeficiente de **lotsize**, 0,00207, parece despreciable y no lo es, porque la parcela se mide en pies cuadrados y va de 1000 a 92681; mil pies cuadrados más de parcela se acompañan de 2,07 mil dólares más. La segunda es el t de **bdrms**, 1,54: datos como estos no serían raros si el coeficiente de **bdrms** valiese cero, es decir, si con la casa y la parcela del mismo tamaño un dormitorio más no se acompañase de ningún cambio de precio. El coeficiente, 13,9 mil dólares, tiene un error estándar de 9,0, y con él la tabla no precisa qué diferencia de precio acompaña a un dormitorio más. La tercera es la constante, -21,8 mil dólares: el valor ajustado de una casa de cero pies cuadrados, cero dormitorios y parcela de cero pies, sin ninguna observación cerca (la casa más pequeña tiene 1171 pies cuadrados): no se lee, y el signo negativo muestra el absurdo.

Tabla 2 - El salario en logaritmos con una indicadora (wage1)

Abra **wage1** y estime el logaritmo del salario por hora sobre educación, experiencia, antigüedad y la indicadora **female**, el modelo de la lección 16.

El **printf** calcula las lecturas exactas, $e^{\beta} - 1$, de la indicadora y de la educación, y el valor $e^{\hat{\beta}_1}$ de la constante en dólares.

```
open wage1 --quiet
```

```
ols lwage const educ exper tenure female
printf "female exacto: %.1f%% ; educ exacto: %.1f%% ; e^const = %.2f dolares ; trabajadores con educ = 0: %d\n",
↪ 100*(exp($coeff(female))-1), 100*(exp($coeff(educ))-1), exp($coeff(const)), sum(educ == 0)
```

Modelo 1: MCO, usando las observaciones 1-526
Variable dependiente: lwage

	coeficiente	Desv. típica	Estadístico t	valor p
const	0,501348	0,101902	4,920	1,16e-06 ***
educ	0,0874623	0,00693890	12,60	5,55e-32 ***
exper	0,00462938	0,00162715	2,845	0,0046 ***
tenure	0,0173670	0,00297622	5,835	9,45e-09 ***
female	-0,301146	0,0372456	-8,085	4,37e-15 ***
Media de la vble. dep.	1,623268	D.T. de la vble. dep.	0,531538	
Suma de cuad. residuos	90,14446	D.T. de la regresión	0,415959	
R-cuadrado	0,392270	R-cuadrado corregido	0,387604	
F(4, 521)	84,07212	Valor p (de F)	4,68e-55	
Log-verosimilitud	-282,4592	Criterio de Akaike	574,9184	
Criterio de Schwarz	596,2449	Crit. de Hannan-Quinn	583,2686	

Log-verosimilitud para wage = -1136,3

female exacto: -26,0% ; educ exacto: 9,1% ; e^{const} = 1,65 dólares ; trabajadores con educ = 0: 2

Escriba: la lectura de **female** en porcentaje, aproximada y exacta, la de **educ**, qué cubre “con lo demás fijo” y qué no cubre, y si la constante se lee.

Lo que debe observar

El modelo es log-nivel: cada coeficiente es una fracción del salario. El de **female**, -0,301, es un 30 % aproximado y un 26 % exacto ($e^{-0,301} - 1 = -0,260$): a una mujer con la misma educación, experiencia y antigüedad que un hombre le corresponde, en promedio, un salario un 26 % menor. La aproximación falla porque 0,30 está por encima del umbral del 10 % de la lección 15. El de **educ**, 0,087, es un 8,7 % aproximado y un 9,1 % exacto por año de educación. “Con lo demás fijo” cubre tres variables, y solo tres: el coeficiente de **female** no dice nada de lo que el modelo no recoge, como la ocupación o el sector, y no es una medida de discriminación. La constante, $e^{0,501} = 1,65$ dólares por hora, es el valor ajustado de un hombre sin educación, experiencia ni antigüedad: solo 2 de los 526 trabajadores tienen **educ** igual a cero, y no se lee.

Tabla 3 - El gasto por alumno y los resultados escolares (meap93)

Abra **meap93** (Wooldridge): 408 institutos de Michigan en 1993, con el porcentaje de alumnos que aprueban la prueba estatal de matemáticas (**math10**), el gasto por alumno en dólares (**expend**), el porcentaje de alumnos en el programa de comedor subvencionado (**lnchprg**, un indicador de pobreza) y el número de alumnos (**enroll**). Estime la tasa de aprobados sobre el logaritmo del gasto, la pobreza y el logaritmo del tamaño.

logs crea **l_expend** y **l_enroll**. El primer **ols** estima el modelo con la pobreza y el tamaño; el primer **printf** traduce el coeficiente del gasto a puntos de aprobados por un 10 % más de gasto y el segundo imprime su intervalo al 95 %, con el valor crítico de la *t* que da **critical** (sesión 18). El segundo **ols** estima la tasa de aprobados sobre el logaritmo del gasto solo, para comparar, y el último **printf** añade la correlación entre el logaritmo del gasto y la pobreza, con la función **corr**.

```
open meap93 --quiet
```

```
summary math10 expend lnchprg enroll --simple
logs expend enroll
ols math10 const l_expend lnchprg l_enroll
printf "10% mas de gasto por alumno: %.2f puntos de aprobados (aprox.)\n", $coeff(l_expend)/10
scalar tc = critical(t, $df, 0.025)
printf "Intervalo al 95% para el coeficiente de l_expend: [%.2f, %.2f]\n", $coeff(l_expend) - tc*$stderr(l_expend),
↪ $coeff(l_expend) + tc*$stderr(l_expend)
```

```

ols math10 const l_expend --quiet
printf "Sin la pobreza ni el tamaño: coeficiente de l_expend %.2f (ee %.2f) ; correlacion entre l_expend y lnchprg: %.3f\n",
↪ $coeff(l_expend), $stderr(l_expend), corr(l_expend, lnchprg)

```

	Media	Mediana	D. T.	Mín	Máx
math10	24,11	23,40	10,49	1,900	66,70
expend	4377	4145	775,8	3332	7419
lnchprg	25,20	23,85	13,61	1,400	79,50
enroll	2664	1840	2697	212,0	16793

Modelo 1: MCO, usando las observaciones 1-408
Variable dependiente: math10

	coeficiente	Desv. típica	Estadístico t	valor p	
const	-23,1377	24,9932	-0,9258	0,3551	
l_expend	7,74606	3,04139	2,547	0,0112	**
lnchprg	-0,323927	0,0363194	-8,919	1,65e-17	***
l_enroll	-1,25544	0,581173	-2,160	0,0313	**
Media de la vble. dep.	24,10686	D.T. de la vble. dep.	10,49361		
Suma de cuad. residuos	36333,69	D.T. de la regresión	9,483400		
R-cuadrado	0,189291	R-cuadrado corregido	0,183271		
F(3, 404)	31,44310	Valor p (de F)	2,75e-18		
Log-verosimilitud	-1494,731	Criterio de Akaike	2997,461		
Criterio de Schwarz	3013,506	Crit. de Hannan-Quinn	3003,810		

10% mas de gasto por alumno: 0,77 puntos de aprobados (aprox.)

Intervalo al 95% para el coeficiente de l_expend: [1,77, 13,72]

Sin la pobreza ni el tamaño: coeficiente de l_expend 11,16 (ee 3,17) ; correlacion entre l_expend y lnchprg: -0,193

Escriba: la lectura del coeficiente del gasto en puntos de aprobados, la de la pobreza, qué dice el R^2 de 0,19 y si la tabla responde a la pregunta “¿mejoran los resultados si se gasta más?”.

Lo que debe observar

El modelo es nivel-log: si el gasto por alumno varía en la fracción p con la pobreza y el tamaño fijos, la tasa de aprobados varía en $7,75p$ puntos; un 10 % más de gasto, unos 0,8 puntos sobre una media de 24. El error estándar es 3,04 y $t = 2,55$: datos como estos serían raros si el coeficiente fuese cero, pero la precisión es baja: el intervalo al 95 % va de 1,8 a 13,7. Un punto más de alumnos en el programa de comedor se acompaña de 0,32 puntos menos de aprobados, con $t = -8,9$. El R^2 es 0,19: el modelo reproduce una quinta parte de la variación entre institutos, y los coeficientes se leen igual; un R^2 bajo no invalida la lectura de un coeficiente, dice que la mayor parte de lo que distingue a los institutos no está en estas tres variables. Sin la pobreza ni el tamaño, el coeficiente del gasto es 11,2: parte de la asociación bruta entre gasto y resultados corresponde a que los institutos con menos pobreza gastan algo más (la correlación entre `l_expend` y `lnchprg` es $-0,19$). Y la pregunta de política no se responde con la tabla: para que 7,75 fuese el efecto de gastar más haría falta que lo que el modelo no recoge fuese ortogonal a las funciones de los regresores, y el gasto por alumno lo deciden distritos que difieren en muchas cosas más.

Tabla 4 - Precios de la vivienda y contaminación, en logaritmos (`hprice2`)

Abra `hprice2`, el fichero de la mayor parte de las prácticas del curso, y estime el logaritmo del precio mediano sobre el logaritmo de la concentración de óxido nítrico, el número de habitaciones, la distancia a los centros de empleo y el número de alumnos por profesor.

`logs` crea `l_price` y `l_nox`. El `printf` da la lectura exacta de `rooms` y de `stratio`.

```
open hprice2 --quiet
```

```
logs price nox
ols l_price const l_nox rooms dist stratio
printf "rooms exacto: %.1f%% ; stratio exacto: %.1f%%\n", 100*(exp($coeff(rooms))-1), 100*(exp($coeff(stratio))-1)
```

Modelo 1: MCO, usando las observaciones 1-506
Variable dependiente: l_price

	coeficiente	Desv. típica	Estadístico t	valor p	
const	11,1521	0,292358	38,15	2,72e-150	***
l_nox	-0,971662	0,101949	-9,531	6,71e-20	***
rooms	0,249067	0,0184996	13,46	1,75e-35	***
dist	-0,0383495	0,00954138	-4,019	6,74e-05	***
stratio	-0,0533867	0,00587202	-9,092	2,28e-18	***

Media de la vble. dep.	9,941057	D.T. de la vble. dep.	0,409255
Suma de cuad. residuos	34,74527	D.T. de la regresión	0,263347
R-cuadrado	0,589213	R-cuadrado corregido	0,585933
F(4, 501)	179,6526	Valor p (de F)	2,42e-95
Log-verosimilitud	-40,32411	Criterio de Akaike	90,64823
Criterio de Schwarz	111,7809	Crit. de Hannan-Quinn	98,93643

rooms exacto: 28,3% ; stratio exacto: -5,2%

Escriba: la lectura de cada coeficiente con su forma funcional (log-log para **nox**, log-nivel para los otros tres), qué unidad de observación es y qué compara el coeficiente de **nox**.

Lo que debe observar

Hay dos lecturas en la misma tabla. **l_nox** es log-log: su coeficiente, $-0,97$ con error estándar $0,10$, es una elasticidad; un 1 % más de concentración se acompaña de un 0,97 % menos de precio, con las habitaciones, la distancia y la ratio de alumnos fijas. Los otros tres son log-nivel: una habitación más de media, un 25 % aproximado y un 28 % exacto más de precio; una milla más de distancia a los centros de empleo, un 3,8 % menos; un alumno más por profesor, un 5,2 % menos. Los cuatro t superan 4 en valor absoluto. La unidad de observación es el barrio, no la vivienda: el coeficiente de **nox** compara precios medianos de barrios con distinta contaminación y las mismas otras tres características. La elasticidad de **nox** es la pregunta de Harrison y Rubinfeld sobre la demanda de aire limpio (lección 15), y la cautela es la de las otras tres tablas: el modelo mide cuánto menos valen las viviendas de los barrios más contaminados, no cuánto subirían si la contaminación bajase.

Preguntas de interpretación para la clase

1. En la Tabla 1, ¿qué coeficiente parece más pequeño y por qué engaña? ¿Qué regla evita el error?
2. En la Tabla 2, ¿por qué el 30 % aproximado y el 26 % exacto difieren tanto, y cuál se informa?
3. En la Tabla 3, el R^2 es 0,19 y en la Tabla 4, 0,59. ¿Qué dice cada uno y qué no dice sobre la lectura de los coeficientes?
4. ¿Cuál de las cuatro constantes se puede leer como cantidad económica? Justifique la respuesta para cada una.
5. Las cuatro tablas provienen de datos observados, no de experimentos. ¿Qué frase es legítima en cada caso y qué frase no lo es?

Para profundizar

- Wooldridge, J. M. (2020), *Introductory Econometrics*, cap. 2, sección 2.4 (unidades de medida y forma funcional), cap. 4, sección 4.2 (los contrastes t y su lectura) y cap. 6, sección 6.1 (efectos de cambiar las unidades).
- Las lecciones 10, 15 y 16 del curso, para las tres formas funcionales y la indicadora.

Código completo de la práctica

```
# -----
# Copyright (C) 2026 Marcos Bujosa
# Licencia: GNU General Public License v3.0 o posterior
# Este código es software libre y puede ser redistribuido y/o modificado bajo los términos de la GPL.
# Ver el archivo LICENSE del repositorio para más detalles.
# -----

# Los dos primeros comandos son necesarios para que Gretl guarde los resultados de la práctica en el
↳ directorio de trabajo
# al ejecutar lo siguiente desde un terminal (use los nombres y ruta que correspondan)
#
#   DIRECTORIO="Nombre_Directorio_trabajo" gretlcli -b ruta/nombre_fichero_de_la_practica.inp
#
# Si esto no le funciona en su sistema, comente las siguientes dos líneas y sitúese en el directorio de
↳ trabajo de gretl
# que corresponda (configure dicho directorio de trabajo desde la ventana principal de Gretl).

string directory = getenv("DIRECTORIO")
set workdir "@directory"

set verbose off

open hprice1 --quiet
outfile --quiet tabla1.txt
    summary price sqrft bdrms lotsize --simple
    ols price const sqrft bdrms lotsize
    printf "100 pies cuadrados mas de casa: %.1f mil dolares ; 1000 pies cuadrados mas de parcela: %.2f mil
↳ dolares\n", 100*$coeff(sqrft), 1000*$coeff(lotsize)
end outfile

open wage1 --quiet
outfile --quiet tabla2.txt
    ols lwage const educ exper tenure female
    printf "female exacto: %.1f%% ; educ exacto: %.1f%% ; e^const = %.2f dolares ; trabajadores con educ =
↳ 0: %d\n", 100*(exp($coeff(female))-1), 100*(exp($coeff(educ))-1), exp($coeff(const)), sum(educ ==
↳ 0)
end outfile

open meap93 --quiet
outfile --quiet tabla3.txt
    summary math10 expend lnchprg enroll --simple
    logs expend enroll
    ols math10 const l_expend lnchprg l_enroll
    printf "10%% mas de gasto por alumno: %.2f puntos de aprobados (aprox.)\n", $coeff(l_expend)/10
    scalar tc = critical(t, $df, 0.025)
    printf "Intervalo al 95%% para el coeficiente de l_expend: [%.2f, %.2f]\n", $coeff(l_expend) -
↳ tc*$stderr(l_expend), $coeff(l_expend) + tc*$stderr(l_expend)
    ols math10 const l_expend --quiet
    printf "Sin la pobreza ni el tamaño: coeficiente de l_expend %.2f (ee %.2f) ; correlacion entre
↳ l_expend y lnchprg: %.3f\n", $coeff(l_expend), $stderr(l_expend), corr(l_expend, lnchprg)
end outfile

open hprice2 --quiet
outfile --quiet tabla4.txt
    logs price nox
    ols l_price const l_nox rooms dist stratio
    printf "rooms exacto: %.1f%% ; stratio exacto: %.1f%%\n", 100*(exp($coeff(rooms))-1),
↳ 100*(exp($coeff(stratio))-1)
end outfile
```

Respuestas

Los cinco párrafos siguientes son los modelos de lectura; las respuestas a las preguntas van después.

Tabla 1. El precio está en miles de dólares. A cien pies cuadrados más de casa, con el mismo número de dormitorios y la misma parcela, les corresponden 12,3 mil dólares más ($t = 9,3$); a mil pies cuadrados más de parcela, con la casa igual, 2,07 mil dólares más ($t = 3,2$). Al número de dormitorios le corresponden 13,9 mil dólares por dormitorio con un error estándar de 9,0 y $t = 1,54$: datos como estos no serían raros si el coeficiente de **bdrms** valiese cero, es decir, si con la superficie y la parcela fijas el dormitorio adicional no se acompañase de ningún cambio de precio. La constante no se interpreta: es el valor ajustado de una casa sin superficie ni parcela, lejos de cualquier dato. El R^2 es 0,67.

Tabla 2. Con la misma educación, experiencia y antigüedad, a una mujer le corresponde, en promedio, un salario un 26 % menor que a un hombre ($e^{-0,301} - 1$; $t = -8,1$). A un año más de educación le acompaña un salario un 9,1 % mayor; a un año más de experiencia, un 0,5 % mayor; a un año más de antigüedad, un 1,7 % mayor, cada uno con los otros tres regresores fijos. El modelo no dice nada de lo que no contiene, y la diferencia entre sexos no se lee como discriminación con estos regresores. La constante no se interpreta: solo dos trabajadores tienen cero años de educación.

Tabla 3. Entre institutos con la misma pobreza y el mismo tamaño, a un 10 % más de gasto por alumno le corresponden 0,8 puntos más de aprobados ($7,75 \times 0,1$), con un error estándar que deja el coeficiente entre 1,8 y 13,7 al 95 %. A un punto más de alumnos en el programa de comedor le corresponden 0,32 puntos menos ($t = -8,9$). El modelo reproduce el 19 % de la variación entre institutos. La tabla no dice que gastar más mejore los resultados: el gasto no se asigna al azar, y lo que el modelo no recoge puede estar correlado con él.

Tabla 4. Entre barrios con el mismo número medio de habitaciones, la misma distancia a los centros de empleo y la misma ratio de alumnos por profesor, a un 1 % más de concentración de óxido nítrico le corresponde un precio mediano un 0,97 % menor ($t = -9,5$). A una habitación más de media, un 28 % más de precio ($e^{0,249} - 1$); a una milla más de distancia, un 3,8 % menos; a un alumno más por profesor, un 5,2 % menos. Se comparan barrios, no viviendas, y la elasticidad mide cuánto menos valen las viviendas donde hay más contaminación, no cuánto subirían si hubiese menos.

1. El de **lotsize**, 0,00207, porque la parcela está en pies cuadrados y los valores son grandes. La regla: leer siempre el coeficiente con la unidad del regresor y, si hace falta, reescalar (mil pies cuadrados, 2,07 mil dólares).
2. Porque $\log(1 + p) \approx p$ solo vale para p pequeño, y 0,30 está muy por encima del 10 % en que la lección 15 situó el límite. Se informa el exacto, 26 %, y se puede decir de qué grupo respecto de cuál.
3. El R^2 mide la fracción de la variación del regresando que el ajuste reproduce: una quinta parte en los institutos, seis décimas en los barrios. No mide la precisión de ningún coeficiente ni su lectura: cada coeficiente se lee con su error estándar. Un R^2 bajo dice que hay mucho que el modelo no recoge. Eso no cambia la lectura *ceteris paribus*, y tampoco decide la cautela causal: esta depende de si lo que el modelo no recoge está relacionado con los regresores, no de cuánto es.
4. Ninguna. En las cuatro tablas el origen de los regresores no constantes, una casa de cero pies, un trabajador con cero años de educación, un instituto con gasto de un dólar, un barrio con una concentración de una unidad y cero habitaciones, queda fuera de los datos. La constante existe y el ajuste la necesita, pero no es una cantidad económica (lección 16).
5. Es legítimo describir cómo varía el regresando entre unidades con distinto valor de un regresor y lo demás fijo: “a cien pies cuadrados más les corresponden 12,3 mil dólares más”. No lo es afirmar lo que pasaría al cambiar el regresor en una casa o un instituto concretos: “ampliar la casa en cien pies sube su precio en 12,3 mil”, “gastar un 10 % más sube los aprobados en 0,8 puntos”. Lo segundo exige el supuesto de exogeneidad de la lección 9, que los datos observados no garantizan.