

Sesión 18 (B) — El ejemplo de la lección 12, en Gretl: regresor concentrado frente a regresor disperso

Marcos Bujosa

Descripción de la práctica

La lección 12 ilustró el intervalo de confianza con una figura de dos paneles: dos ajustes con *la misma* muestra de doce perturbaciones, los mismos parámetros verdaderos y el mismo residuo, en los que lo único distinto era cuánto se separaban los valores del regresor. En uno el intervalo era $[0,08; 0,90]$; en el otro, $[0,46; 0,54]$.

Esta práctica rehace esa figura en Gretl, con sus 24 puntos exactos. No es un ejemplo nuevo: es el de la lección, para que pueda reproducir cada número que allí se dio por bueno y ver de dónde sale.

Los datos son *simulados con parámetros conocidos*: $\beta_1 = 10$, $\beta_2 = 0,5$, y doce perturbaciones normales de desviación típica en torno a 1,1. Conocer la verdad es lo que permite, en la Actividad 3, contrastar una hipótesis que *sabemos* cierta —cosa imposible con datos reales— y preparar así la práctica C.

Objetivos

1. Reproducir los dos ajustes de la figura de la lección 12 y sus dos intervalos de confianza.
2. Verificar que $ee(\hat{\beta}_2) = s/\|x - \bar{x}\|_e$ y que, con el mismo residuo, toda la diferencia de precisión está en la longitud del regresor.
3. Contrastar $H_0: \beta_2 = 0,5$ —que es cierta— en los dos ajustes.
4. Ver qué columnas de la tabla cambian, y cuáles no, al cambiar las unidades del regresor.

Actividad 1 - Regresor concentrado

Los doce valores del regresor van de 47,25 a 52,75 en pasos de 0,5; los del regresando son los del panel izquierdo de la figura.

en línea de comandos:

```
nulldata 12
setobs 1 1 --cross-section

series x_c = 47.25 + 0.5*(index - 1)
series y_c = {35.045, 33.025, 34.435, 32.415, 35.365, 35.975, \
              34.655, 34.055, 36.505, 35.925, 35.495, 37.105}'

ols y_c const x_c
scalar tc = critical(t, $df, 0.025)
printf "\n||x - xbar|| = %.4f    s = %.4f    ee = s/||x - xbar|| = %.4f (Gretl: %.4f)\n", sqrt(sum((x_c-mean(x_c))^2)),
↪ $sigma, $sigma/sqrt(sum((x_c-mean(x_c))^2)), $stderr(x_c)
printf "t = %.2f    (t_c al 5% = %.3f)\n", $coeff(x_c)/$stderr(x_c), tc
printf "IC 95%: [%.2f ; %.2f]\n", $coeff(x_c) - tc*$stderr(x_c), $coeff(x_c) + tc*$stderr(x_c)
```

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

Modelo 1: MCO, usando las observaciones 1-12
Variable dependiente: y_c

	coeficiente	Desv. típica	Estadístico t	valor p
const	10,3427	9,20510	1,124	0,2874
x_c	0,493147	0,183992	2,680	0,0231 **
Media de la vble. dep.	35,00000	D.T. de la vble. dep.	1,374994	
Suma de cuad. residuos	12,10252	D.T. de la regresión	1,100115	
R-cuadrado	0,418056	R-cuadrado corregido	0,359861	
F(1, 10)	7,183775	Valor p (de F)	0,023083	
Log-verosimilitud	-17,07831	Criterio de Akaike	38,15661	
Criterio de Schwarz	39,12642	Crit. de Hannan-Quinn	37,79755	

$|x - \bar{x}| = 5,9791$ $s = 1,1001$ $ee = s/|x - \bar{x}| = 0,1840$ (Gretl: 0,1840)
 $t = 2,68$ (t_c al 5% = 2,228)
 IC 95%: [0,08 ; 0,90]

Lo que debe observar

Son los números del panel izquierdo de la lección: $\hat{\beta}_2 = 0,4931$, error estándar 0,184, $|t| = 2,68$ e intervalo [0,08; 0,90]. El coeficiente es significativo al 5% —el cero queda fuera— y el intervalo es tan ancho que no dice casi nada sobre cuánto vale la pendiente. Fíjese también en el error estándar reconstruido: s dividido por una longitud de solo 5,98.

Actividad 2 - Regresor disperso

Ahora los valores del regresor van de 22,5 a 77,5 en pasos de 5 —diez veces más separados, alrededor del mismo centro—, y el regresando es el del panel derecho.

en línea de comandos:

```
series x_d = 22.5 + 5*(index - 1)
series y_d = {22.670, 22.900, 26.560, 26.790, 31.990, 34.850, \
              35.780, 37.430, 42.130, 43.800, 45.620, 49.480}'

ols y_d const x_d
printf "\n|x - xbar| = %.4f    s = %.4f    ee = s/|x - xbar| = %.4f    (Gretl: %.4f)\n", sqrt(sum((x_d-mean(x_d))^2)),
↪ $sigma, $sigma/sqrt(sum((x_d-mean(x_d))^2)), $stderr(x_d)
printf "t = %.2f    (t_c al 5% = %.3f)\n", $coeff(x_d)/$stderr(x_d), tc
printf "IC 95%: [%.2f ; %.2f]\n", $coeff(x_d) - tc*$stderr(x_d), $coeff(x_d) + tc*$stderr(x_d)

series u_c = y_c - 10 - 0.5*x_c                    # perturbacion del panel concentrado
series u_d = y_d - 10 - 0.5*x_d                    # perturbacion del panel disperso
printf "\nmayor diferencia entre las dos series de perturbaciones: %.6f\n", max(abs(u_c - u_d))
```

Modelo 2: MCO, usando las observaciones 1-12
Variable dependiente: y_d

	coeficiente	Desv. típica	Estadístico t	valor p
const	10,0343	0,973234	10,31	1,20e-06 ***
x_d	0,499315	0,0183992	27,14	1,07e-10 ***
Media de la vble. dep.	35,00000	D.T. de la vble. dep.	9,062431	
Suma de cuad. residuos	12,10252	D.T. de la regresión	1,100115	
R-cuadrado	0,986603	R-cuadrado corregido	0,985264	
F(1, 10)	736,4595	Valor p (de F)	1,07e-10	

Log-verosimilitud	-17,07831	Criterio de Akaike	38,15661
Criterio de Schwarz	39,12642	Crit. de Hannan-Quinn	37,79755

```
||x - xbar|| = 59,7913    s = 1,1001    ee = s/||x - xbar|| = 0,0184 (Gretl: 0,0184)
t = 27,14    (t_c al 5% = 2,228)
IC 95%: [0,46 ; 0,54]
```

mayor diferencia entre las dos series de perturbaciones: 0,000000

Lo que debe observar

La s es la misma que en la Actividad 1, 1,1001, hasta la cuarta cifra. Y la última línea lo explica: las doce perturbaciones —lo que queda al restar de cada y_i la parte verdadera $10 + 0,5x_i$ — son idénticas en los dos paneles. Como el regresor en desviaciones apunta en la misma dirección en ambos casos (es el mismo patrón, multiplicado por diez), la proyección del ruido sobre esa dirección, y con ella el residuo, coincide.

Toda la diferencia está, por tanto, en el denominador del error estándar: la longitud del regresor en desviaciones es 59,79 en vez de 5,98. Diez veces más largo el regresor, diez veces más pequeño el error estándar (0,018 frente a 0,184), diez veces más estrecho el intervalo ([0,46; 0,54]). El ruido no ha cambiado; lo que ha cambiado es cuánto varía la variable cuyo efecto queremos medir. *Para estimar bien el efecto de una variable hace falta que esa variable varíe.*

Actividad 3 - Contrastar la hipótesis verdadera

Con datos reales no sabemos cuánto vale β_2 . Aquí sí: vale 0,5, porque lo hemos elegido nosotros. Podemos entonces contrastar $H_0 : \beta_2 = 0,5$ sabiendo que es *cierta*, y ver qué hace el contraste con ella; y de paso, contrastar $H_0 : \beta_2 = 0$, que sabemos *falsa*.

en línea de comandos:

```
ols y_c const x_c --quiet
scalar t_cierta_c = ($coeff(x_c) - 0.5)/$stderr(x_c)
scalar t_falsa_c = ($coeff(x_c) - 0.0)/$stderr(x_c)
ols y_d const x_d --quiet
scalar t_cierta_d = ($coeff(x_d) - 0.5)/$stderr(x_d)
scalar t_falsa_d = ($coeff(x_d) - 0.0)/$stderr(x_d)

printf "t_c al 5% con %d grados de libertad = %.3f\n", $df, tc
printf "          concentrado      disperso\n"
printf "H0: beta_2 = 0,5 (cierta)  t = %6.2f    t = %6.2f\n", t_cierta_c, t_cierta_d
printf "H0: beta_2 = 0   (falsa)   t = %6.2f    t = %6.2f\n", t_falsa_c, t_falsa_d
```

t_c al 5% con 10 grados de libertad = 2,228

	concentrado	disperso
H0: beta_2 = 0,5 (cierta)	t = -0,04	t = -0,04
H0: beta_2 = 0 (falsa)	t = 2,68	t = 27,14

Lo que debe observar

La hipótesis cierta no se rechaza en ningún panel —los dos estadísticos están cerca de cero—, y el 0,5 verdadero está dentro de los dos intervalos: dentro del ancho y dentro del estrecho. La hipótesis falsa se rechaza en los dos, pero con evidencias muy desiguales: 2,68 frente a 27,14.

Aquí conviene resistir una tentación. Que el intervalo contenga al verdadero β_2 *esta vez* no es la garantía del intervalo: la garantía, según la lección 12, es sobre el procedimiento —“el 95% de las muestras dan un intervalo que contiene a β_2 ”—, y con una sola muestra no se puede comprobar una frecuencia. Para eso hay que repetir el experimento muchas veces, con nuevas perturbaciones cada vez, y contar. Eso es la práctica C.

Actividad 4 - Cambiar las unidades del regresor

Midamos el regresor disperso en centésimas (multiplicamos x_d por cien) y reestimemos.
en línea de comandos:

```
series x_d100 = 100*x_d
ols y_d const x_d --quiet
printf "x en unidades originales : coef = %.6f ee = %.6f t = %.2f p = %.3g\n", $coeff(x_d), $stderr(x_d),
↪ $coeff(x_d)/$stderr(x_d), 2*pvalue(t, $df, abs($coeff(x_d)/$stderr(x_d)))
ols y_d const x_d100 --quiet
printf "x multiplicado por 100 : coef = %.6f ee = %.6f t = %.2f p = %.3g\n", $coeff(x_d100), $stderr(x_d100),
↪ $coeff(x_d100)/$stderr(x_d100), 2*pvalue(t, $df, abs($coeff(x_d100)/$stderr(x_d100)))
printf "R2 en ambos casos : %.4f\n", $rsq
```

```
x en unidades originales : coef = 0,499315 ee = 0,018399 t = 27,14 p = 1,07e-10
x multiplicado por 100 : coef = 0,004993 ee = 0,000184 t = 27,14 p = 1,07e-10
R2 en ambos casos : 0,9866
```

Lo que debe observar

El coeficiente y su error estándar se dividen los dos por cien; el estadístico t , el valor p y el R^2 no se mueven. Es la pregunta 10 de la lección 12 con números: las dos primeras columnas de la tabla están en las unidades del problema y las dos últimas son adimensionales. Por eso t y p sirven para comparar y no sirven para valorar la magnitud de un efecto, y por eso el coeficiente hay que leerlo siempre sabiendo en qué unidades está.

Preguntas de interpretación para la clase

1. Los dos paneles tienen el mismo residuo. ¿Por qué? ¿Seguiría siendo así si, en lugar de multiplicar por diez los valores del regresor, los hubiéramos sorteado de nuevo?
2. En la Actividad 1, el intervalo $[0,08; 0,90]$ no contiene al cero. ¿Diría usted que ese ajuste “demuestra” que la pendiente es positiva? ¿Y que “demuestra” que vale 0,5?
3. Los dos ajustes tienen el mismo s y errores estándar que difieren en un factor diez. Escriba la fórmula que lo explica y diga qué cantidad de la lección 4 aparece en ella.
4. En la Actividad 3 la hipótesis cierta no se rechaza. ¿Podría haberse rechazado? ¿Con qué frecuencia cabría esperar que eso ocurriera si repitiéramos el experimento con nuevas perturbaciones?
5. Tras la Actividad 4, ¿qué columnas de una tabla de resultados usaría para comparar la *precisión relativa* con que se han estimado dos coeficientes medidos en unidades distintas? ¿Y para comparar sus *magnitudes*?

Para profundizar

- Lección 12 ([lección 12](#)), secciones “El error estándar, y el precio de estimar σ ” y “El intervalo de confianza”, cuya figura se ha reproducido aquí.
- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 2, sección 2.5 (varianza de los estimadores MCO: el papel de la variación del regresor) y cap. 4, sección 4.3.

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

nulldata 12
setobs 1 1 --cross-section

series x_c = 47.25 + 0.5*(index - 1)
series y_c = {35.045, 33.025, 34.435, 32.415, 35.365, 35.975, \
              34.655, 34.055, 36.505, 35.925, 35.495, 37.105}'
outfile --quiet concentrado.txt
  ols y_c const x_c
  scalar tc = critical(t, $df, 0.025)
  printf "\n||x - xbar|| = %.4f    s = %.4f    ee = s/||x - xbar|| = %.4f (Gretl: %.4f)\n",
  ↪ sqrt(sum((x_c-mean(x_c))^2)), $sigma, $sigma/sqrt(sum((x_c-mean(x_c))^2)), $stderr(x_c)
  printf "t = %.2f    (t_c al 5%% = %.3f)\n", $coeff(x_c)/$stderr(x_c), tc
  printf "IC 95%%: [%.2f ; %.2f]\n", $coeff(x_c) - tc*$stderr(x_c), $coeff(x_c) + tc*$stderr(x_c)
end outfile

series x_d = 22.5 + 5*(index - 1)
series y_d = {22.670, 22.900, 26.560, 26.790, 31.990, 34.850, \
              35.780, 37.430, 42.130, 43.800, 45.620, 49.480}'
outfile --quiet disperso.txt
  ols y_d const x_d
  printf "\n||x - xbar|| = %.4f    s = %.4f    ee = s/||x - xbar|| = %.4f (Gretl: %.4f)\n",
  ↪ sqrt(sum((x_d-mean(x_d))^2)), $sigma, $sigma/sqrt(sum((x_d-mean(x_d))^2)), $stderr(x_d)
  printf "t = %.2f    (t_c al 5%% = %.3f)\n", $coeff(x_d)/$stderr(x_d), tc
  printf "IC 95%%: [%.2f ; %.2f]\n", $coeff(x_d) - tc*$stderr(x_d), $coeff(x_d) + tc*$stderr(x_d)
  series u_c = y_c - 10 - 0.5*x_c          # perturbacion del panel concentrado
  series u_d = y_d - 10 - 0.5*x_d          # perturbacion del panel disperso
  printf "\nmayor diferencia entre las dos series de perturbaciones: %.6f\n", max(abs(u_c - u_d))
end outfile

outfile --quiet contrastes.txt
  ols y_c const x_c --quiet
  scalar t_cierta_c = ($coeff(x_c) - 0.5)/$stderr(x_c)
  scalar t_falsa_c = ($coeff(x_c) - 0.0)/$stderr(x_c)
  ols y_d const x_d --quiet
  scalar t_cierta_d = ($coeff(x_d) - 0.5)/$stderr(x_d)
  scalar t_falsa_d = ($coeff(x_d) - 0.0)/$stderr(x_d)

  printf "t_c al 5%% con %d grados de libertad = %.3f\n\n", $df, tc
  printf "                                concentrado    disperso\n"
  printf "H0: beta_2 = 0,5 (cierta)    t = %6.2f    t = %6.2f\n", t_cierta_c, t_cierta_d
  printf "H0: beta_2 = 0    (falsa)    t = %6.2f    t = %6.2f\n", t_falsa_c, t_falsa_d
end outfile
```

```

outfile --quiet unidades.txt
series x_d100 = 100*x_d
ols y_d const x_d --quiet
printf "x en unidades originales : coef = %.6f   ee = %.6f   t = %.2f   p = %.3g\n", $coeff(x_d),
↪ $stderr(x_d), $coeff(x_d)/$stderr(x_d), 2*pvalue(t, $df, abs($coeff(x_d)/$stderr(x_d)))
ols y_d const x_d100 --quiet
printf "x multiplicado por 100 : coef = %.6f   ee = %.6f   t = %.2f   p = %.3g\n", $coeff(x_d100),
↪ $stderr(x_d100), $coeff(x_d100)/$stderr(x_d100), 2*pvalue(t, $df,
↪ abs($coeff(x_d100)/$stderr(x_d100)))
printf "R2 en ambos casos : %.4f\n", $rsq
end outfile

```

Enlace al guión: [S18-Prct-B-regresor-disperso.inp](#)

Respuestas

1. *El mismo residuo.* El residuo es lo que queda de $\mathbf{y}$ tras quitarle su proyección sobre $\mathcal{L}(\mathbf{1}, \mathbf{x})$ (lección 6), y ese subespacio es el mismo en los dos paneles: multiplicar $\mathbf{x}$ por diez no cambia el plano que $\mathbf{1}$ y $\mathbf{x}$ generan, solo el nombre de sus generadores. Como además $\mathbf{y}$ se construyó con las mismas perturbaciones, la parte de $\mathbf{y}$ ortogonal a ese plano —el residuo— es idéntica. Si hubiéramos sorteado de nuevo los valores del regresor, el plano sería otro y el residuo, en general, distinto.
2. *Qué “demuestra” el intervalo ancho.* Ni una cosa ni la otra, en sentido estricto. Que el cero quede fuera dice que los datos son poco compatibles con una pendiente nula, con un 5 % de riesgo de equivocarse: no es una demostración, es un veredicto probabilístico. Y sobre el valor 0,5 no dice nada especial: el intervalo lo contiene, pero contiene también al 0,2 y al 0,8. Un intervalo ancho es compatible con muchas verdades distintas, y esa es su información: que los datos no permiten precisar más.
3. *La fórmula.* $ee(\hat{\beta}_2) = \mathbf{s}/\|\mathbf{x} - \bar{\mathbf{x}}\|_e$. El denominador es la norma del vector en desviaciones del regresor, el objeto que la lección 4 definió al construir la desviación típica (allí con el producto escalar estadístico; aquí con el euclídeo, sin el factor $1/n$, porque así es como aparece en la varianza de $\hat{\beta}_2$ de la lección 11). Con el mismo $\mathbf{s}$ en el numerador, dos regresores cuyas longitudes en desviaciones están en razón 1 : 10 dan errores estándar en razón 10 : 1.
4. *¿Podría haberse rechazado la hipótesis cierta?* Sí. El contraste al 5 % rechaza una hipótesis cierta, por construcción, el 5 % de las veces: es lo que significa el nivel de significación. En esta muestra concreta no ha ocurrido —los estadísticos son $-0,04$ —, pero en una de cada veinte muestras del mismo mundo el 0,5 quedaría fuera del intervalo. Esa frecuencia no se puede ver con una muestra; se ve repitiendo, y eso hace la práctica C.
5. *Precisión relativa y magnitud.* Para la precisión relativa, el estadístico t (o el valor p), porque son adimensionales: dicen a cuántos errores estándar está cada coeficiente de cero, sea cual sea su unidad. Para la magnitud, solo sirve el coeficiente en las unidades del problema —y para compararla entre variables medidas en unidades distintas hay que reexpresarlas en una escala común (por ejemplo, el efecto de un cambio de una desviación típica en cada regresor), cosa que las columnas de la tabla no hacen por sí solas.