

Sesión 18 (A) — Leer una tabla de resultados: contrastes t e intervalos con datos reales

Marcos Bujosa

Descripción de la práctica

La lección 12 construyó tres herramientas —el error estándar, el intervalo de confianza y el contraste t — y dedicó su segunda mitad a algo que allí se dijo que usted usaría más veces que cualquier otra cosa del curso: *leer una tabla de resultados*. Hoy se trata de hacerlo con datos reales y con el programa delante.

Retomamos `hprice2` y la regresión `price ~ rooms`, el ajuste que arrastramos desde la práctica que siguió a la lección 8 y que sirvió de “mundo verdadero” en el laboratorio anterior. Es la última vez que ese ajuste nos sirve tal cual: hoy, además de estimarlo, le pedimos que hable de lo que *no* sabemos.

Trabajamos con datos *reales* porque el objetivo de esta práctica es la *lectura*: aprender a distinguir, en una salida de Gretl en español, qué número responde a qué pregunta. La verificación con datos simulados —¿cubre de verdad el intervalo el 95\

Objetivos

1. Leer las cuatro columnas de una tabla de resultados e identificar, en la salida en español, las *tres* cantidades que Gretl rotula como “desviación típica”.
2. Reconstruir a mano, con los accesorios de Gretl, el error estándar, el estadístico t , el valor p y el intervalo de confianza, y comprobar que coinciden con la tabla.
3. Verificar sobre datos reales la identidad de la lección 12, $t^2 = (n - k) \text{SEC}/\text{SRC}$, y reconocer en ella el F que Gretl imprime al pie de la tabla.
4. Contrastar una hipótesis distinta de “el coeficiente es cero” y ver que el intervalo de confianza ya contenía la respuesta.
5. Repetir la estimación con doce viviendas para ver qué cambia —y qué no— cuando la evidencia es escasa.

Actividad 1 - La tabla de resultados y sus cuatro columnas

Abra `hprice2` y estime la regresión de `price` sobre `rooms`: en la ventana principal, Modelo -> Mínimos cuadrados ordinarios, con `price` como variable dependiente y `rooms` como regresor.

o bien teclee en línea de comandos:

```
open hprice2.gdt --quiet
ols price const rooms
```

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

coeficiente	Desv. típica	Estadístico t	valor p
-------------	--------------	---------------	---------

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

const	-34796,2	2651,53	-13,12	4,74e-34 ***
rooms	9119,55	419,339	21,75	1,87e-74 ***
Media de la vble. dep.	22511,51	D.T. de la vble. dep.	9208,856	
Suma de cuad. residuos	2,21e+10	D.T. de la regresión	6620,865	
R-cuadrado	0,484110	R-cuadrado corregido	0,483086	
F(1, 504)	472,9521	Valor p (de F)	1,87e-74	
Log-verosimilitud	-5168,759	Criterio de Akaike	10341,52	
Criterio de Schwarz	10349,97	Crit. de Hannan-Quinn	10344,83	

Lo que debe observar

La tabla tiene, para cada regresor, las cuatro columnas de la lección 12, y conviene leerlas en el orden en que responden preguntas:

- *Coficiente*: el tamaño del efecto, en las unidades del problema. Aquí, $\hat{\beta}_2 = 9\,119,55$: en el ajuste, a una habitación más le corresponden unos 9 120 dólares más de precio mediano.
- *Desv. típica*: el error estándar $ee(\hat{\beta}_2) = 419,34$. Es la columna que dice con cuánta precisión se ha medido el coeficiente, y es la que se ignora con más frecuencia.
- *Estadístico t*: el cociente de las dos anteriores, $9\,119,55/419,34 = 21,75$. Cuántos errores estándar separan al coeficiente de cero.
- *Valor p*: la probabilidad de un $|t|$ al menos así de grande si el coeficiente fuese cero. Aquí es del orden de 10^{-74} : no hay ninguna duda de que **rooms** interviene.

Y ahora el aviso terminológico de la lección, que en la salida en español es inevitable. Busque en la tabla la palabra “típica” o la abreviatura “D.T.”: aparece **tres veces**, para tres cantidades que no tienen nada que ver entre sí.

Etiqueta en la salida	Qué es	Valor aquí
D.T. de la vble. dep.	dispersión de los <i>datos</i> de <i>price</i> : $\sqrt{STC/(n-1)}$	9 208,9
D.T. de la regresión	$s = \sqrt{SRC/(n-k)}$: estimación de la desviación típica de la <i>perturbación</i>	6 620,9
Desv. típica (de rooms)	$ee(\hat{\beta}_2)$: dispersión del <i>estimador</i> de una muestra a otra	419,3

En inglés solo la primera es *standard deviation*; las otras dos son *standard error*. Al leer una salida en español, identifíquelas por su posición, no por su nombre.

Actividad 2 - La tabla, reconstruida a mano

Ninguno de los números de la tabla es un misterio: todos salen de las fórmulas de la lección 12 y de cantidades que ya sabíamos calcular. Vamos a rehacerlos uno a uno con los *accesores* de Gretl (los objetos que empiezan por \$ y que conocemos desde la práctica que siguió a la lección 8) y a comprobar que coinciden.
en línea de comandos:

```

scalar n = $nobs
scalar k = 2
scalar SRC = $ess
scalar STC = SRC/(1 - $rsq)
scalar SEC = STC - SRC
scalar s = sqrt(SRC/(n-k))
scalar Lx = sqrt(sum((rooms - mean(rooms))^2))
scalar ee = s/Lx
scalar b2 = $coeff(rooms)
scalar tst = b2/ee
scalar tc = critical(t, n-k, 0.025)
scalar p = 2*pvalue(t, n-k, abs(tst))

# regresores, constante incluida
# suma de cuadrados de los residuos
# Gretl no da STC directamente; R^2 = 1 - SRC/STC
# raíz de la cuasivarianza de los residuos
# longitud euclídea de x - xbar
# error estandar: ruido por unidad de longitud
# valor critico al 5%, dos colas
# valor p bilateral

```

```
printf "n - k                = %d\n", n-k
printf "s (D.T. de la regresion) = %.4f   Gretl: %.4f\n", s, $sigma
printf "||x - xbar||         = %.4f\n", Lx
printf "ee(b2) = s/||x - xbar|| = %.4f   Gretl: %.4f\n", ee, $stderr.rooms)
printf "t = b2/ee            = %.4f\n", tst
printf "valor critico t_c (5%) = %.4f\n", tc
printf "valor p              = %.3g\n", p
printf "IC 95%: [%.2f ; %.2f]\n", b2 - tc*ee, b2 + tc*ee
```

```
n - k                = 504
s (D.T. de la regresion) = 6620,8652   Gretl: 6620,8652
||x - xbar||         = 15,7888
ee(b2) = s/||x - xbar|| = 419,3385   Gretl: 419,3385
t = b2/ee            = 21,7475
valor critico t_c (5%) = 1,9647
valor p              = 1,87e-74
IC 95%: [8295,68 ; 9943,42]
```

Lo que debe observar

Cada línea reproduce un número de la tabla. Fíjese en particular en el error estándar: es, literalmente, la s de la lección 11 dividida por la longitud del regresor en desviaciones, $\|\mathbf{x} - \bar{\mathbf{x}}\|_e$, que aquí hemos calculado tal como se definió en la lección 4. Es el “ruido por unidad de longitud del regresor” de la lección 12, y no hay nada más dentro.

Dos detalles de la reconstrucción:

- Gretl no ofrece un accesor para STC, pero sí para SRC (`$ess`) y para R^2 (`$rsq`); como $R^2 = 1 - \text{SRC}/\text{STC}$, basta despejar.
- `critical(t, gl, 0.025)` devuelve el valor que deja una probabilidad 0,025 en la cola derecha de la t con `gl` grados de libertad —de ahí que el 5% bilateral se pida como 0,025—; y `pvalue(t, gl, x)` devuelve la probabilidad a la derecha de x , por lo que el valor p bilateral se obtiene multiplicando por dos la de $|t|$.

También puede obtener el intervalo desde los menús: en la ventana del modelo estimado, **Análisis** -> **Intervalos de confianza para los coeficientes**. Compruebe que coincide con el calculado.

Actividad 3 - La identidad de la lección 12, y el F del pie de la tabla

La lección 12 demostró que el estadístico t no contiene nada que no estuviera ya en el triángulo de la lección 8:

$$t^2 = (n - k) \frac{\text{SEC}}{\text{SRC}} = (n - k) \frac{R^2}{1 - R^2}.$$

Comprobémoslo con los tres números que acabamos de calcular.

en línea de comandos:

```
printf "sqrt((n-k)*SEC/SRC) = %.4f   t = %.4f\n", sqrt((n-k)*SEC/SRC), tst
printf "(n-k)*R2/(1-R2)     = %.4f   t^2 = %.4f\n", (n-k)*$rsq/(1-$rsq), tst^2
printf "F(1, %d) del pie de la tabla = %.4f\n", n-k, $Fstat
```

```
sqrt((n-k)*SEC/SRC) = 21,7475   t = 21,7475
(n-k)*R2/(1-R2)     = 472,9521   t^2 = 472,9521
F(1, 504) del pie de la tabla = 472,9521
```

Lo que debe observar

Las dos vías dan el mismo t , y t^2 coincide con el $F(1, 504)$ que Gretl imprime al pie de toda tabla de regresión. No es coincidencia: ese F es el cociente entre la parte explicada y la parte residual, cada una dividida por su número de dimensiones, y con un solo regresor no constante es exactamente t^2 . La lección 13 lo convertirá en un contraste por derecho propio, capaz de juzgar varios regresores a la vez; hoy basta con haberlo reconocido.

Puede verlo también desde los menús: en la ventana del modelo, **Contrastes -> Restricciones lineales**, escriba $b[2] = 0$ y acepte. Gretl responde con un estadístico F y su valor p ; compárelos con t^2 y con el valor p de la tabla.

o bien teclee en línea de comandos:

```
restrict
  b[2] = 0
end restrict
```

Restricción:
 $b[\text{rooms}] = 0$

Estadístico de contraste: $F(1, 504) = 472,952$, con valor $p = 1,86879\text{e-}74$

Estimaciones restringidas:

	coeficiente	Desv. típica	Estadístico t	valor p
const	22511,5	409,384	54,99	2,45e-215 ***
rooms	0,00000	0,00000	NA	NA

Desviación típica de la regresión = 9208,86

Actividad 4 - Contrastar contra un valor distinto de cero

Que **rooms** intervenga no es ninguna noticia. Una pregunta con más sustancia es, por ejemplo: *¿es compatible con estos datos que una habitación valga diez mil dólares?* Es decir, $H_0: \beta_2 = 10000$ frente a $H_1: \beta_2 \neq 10000$.

en línea de comandos:

```
scalar c      = 10000
scalar t_c0 = (b2 - c)/ee
scalar p_c0 = 2*pvalue(t, n-k, abs(t_c0))
scalar tc1 = critical(t, n-k, 0.005)      # valor critico al 1%, dos colas

printf "H0: beta_2 = %d\n", c
printf "t = (b2 - c)/ee      = %.4f\n", t_c0
printf "valor p              = %.4f\n", p_c0
printf "t_c al 5%            = %.4f  -> %s\n", tc, (abs(t_c0) > tc ? "se rechaza" : "no se rechaza")
printf "t_c al 1%            = %.4f  -> %s\n", tc1, (abs(t_c0) > tc1 ? "se rechaza" : "no se rechaza")
printf "IC 95%: [%.2f ; %.2f] ¿contiene a %d? %s\n", b2 - tc*ee, b2 + tc*ee, c, (b2 - tc*ee <= c && c <= b2 + tc*ee
↵ ? "si" : "no")
printf "IC 99%: [%.2f ; %.2f] ¿contiene a %d? %s\n", b2 - tc1*ee, b2 + tc1*ee, c, (b2 - tc1*ee <= c && c <= b2 + tc1*ee
↵ ? "si" : "no")
```

```
H0: beta_2 = 10000
t = (b2 - c)/ee      = -2,0996
valor p              = 0,0363
t_c al 5%            = 1,9647  -> se rechaza
t_c al 1%            = 2,5856  -> no se rechaza
IC 95%: [8295,68 ; 9943,42] ¿contiene a 10000? no
IC 99%: [8035,30 ; 10203,80] ¿contiene a 10000? si
```

Lo que debe observar

El estadístico vale $-2,10$: la estimación está algo más de dos errores estándar *por debajo* de 10 000. Al 5 % eso basta para rechazar ($|t| > 1,96$); al 1 % no ($|t| < 2,59$). El veredicto depende del nivel, y por eso el valor $p = 0,036$ informa más que cualquiera de los dos veredictos.

Ahora mire los intervalos: el de 95 % deja fuera a 10 000 y el de 99 % lo contiene. Es la equivalencia exacta de la lección 12: un valor queda fuera del intervalo al $1 - \alpha$ si y solo si el contraste contra ese valor se rechaza al nivel α . El intervalo ya contenía la respuesta a esta pregunta —y a cualquier otra de la forma “¿es compatible con $\beta_2 = c$?”— antes de que la formuláramos. Pruebe en clase otros valores de c : el intervalo le dirá el resultado antes de calcular nada.

Actividad 5 - La misma regresión con doce viviendas

Todo lo anterior descansa sobre 506 observaciones. ¿Qué cambia si solo tuviéramos una docena? Tomamos doce viviendas al azar¹ y repetimos.

en línea de comandos:

```
set seed 20261107                                # para que todos obtengamos las mismas doce
smpl 12 --random
ols price const rooms
scalar tc12 = critical(t, $df, 0.025)
printf "\nn = %d, grados de libertad = %d, t_c al 5%% = %.3f (con 506 viviendas era %.3f)\n", $nobs, $df, tc12, tc
printf "IC 95% con 12 viviendas : [%.1f ; %.1f]\n", $coeff(rooms) - tc12*$stderr(rooms), $coeff(rooms) +
  ↪ tc12*$stderr(rooms)
printf "IC 95% con 506 viviendas: [%.1f ; %.1f]\n", b2 - tc*ee, b2 + tc*ee
smpl full
```

Modelo 2: MCO, usando las observaciones 1-12

Variable dependiente: price

	coeficiente	Desv. típica	Estadístico t	valor p
const	-46925,4	25791,1	-1,819	0,0989 *
rooms	10927,6	4056,86	2,694	0,0226 **
Media de la vble. dep.	22291,75	D.T. de la vble. dep.	9560,430	
Suma de cuad. residuos	5,83e+08	D.T. de la regresión	7633,254	
R-cuadrado	0,420475	R-cuadrado corregido	0,362523	
F(1, 10)	7,255519	Valor p (de F)	0,022560	
Log-verosimilitud	-123,2166	Criterio de Akaike	250,4331	
Criterio de Schwarz	251,4029	Crit. de Hannan-Quinn	250,0741	

$n = 12$, grados de libertad = 10, t_c al 5% = 2,228 (con 506 viviendas era 1,965)

IC 95% con 12 viviendas : [1888,3 ; 19966,8]

IC 95% con 506 viviendas: [8295,7 ; 9943,4]

Lo que debe observar

Con estas doce viviendas el coeficiente sigue siendo *significativo*: $|t| = 2,69$ frente a un valor crítico de 2,23. Si solo mirase el asterisco, concluiría lo mismo que con quinientas. Pero mire el intervalo: va de unos 1 900 a casi 20 000 dólares por habitación, diez veces más ancho que el de la muestra completa. Es exactamente la situación que la lección 12 describió con el panel izquierdo de su figura: *significativo, e informativo de casi nada*.

Dos cosas han cambiado y una no. Han cambiado el error estándar (diez veces mayor, porque el regresor en desviaciones es mucho más corto con doce datos) y el valor crítico (2,23 con diez grados de libertad frente

¹Al azar es esencial: `hprice2` viene ordenado por distritos, de modo que las doce primeras filas son vecinas entre sí y anómalamente parecidas. Con `smpl 1 12` el experimento engañaría, como ya se advirtió en el laboratorio anterior; `smpl 12 --random` elige doce filas cualesquiera.

a 1,96 con quinientos: la t exige más cuando hay menos evidencia). No ha cambiado la relación entre las variables, que es la misma en la ciudad de Boston se muestreen las casas que se muestreen. Por eso la lección insistía en mirar el coeficiente y el intervalo antes que el veredicto.

Preguntas de interpretación para la clase

1. En la salida de la Actividad 1 aparecen tres cantidades rotuladas como “desviación típica”. ¿Cuál de ellas cambiaría si dispusiéramos de más viviendas con la misma relación entre precio y habitaciones, y cuáles no? Razónelo con las fórmulas.
2. En la Actividad 2 el error estándar resultó ser $s/\|\mathbf{x} - \bar{\mathbf{x}}\|_e$. Si en lugar de `rooms` usáramos `rooms` medido en decenas de habitaciones, ¿qué le pasaría a cada uno de los dos factores? ¿Y al estadístico t ?
3. La Actividad 3 mostró que t^2 coincide con el F del pie de la tabla. ¿Qué hay en ese F que no estuviera ya en el R^2 ?
4. En la Actividad 4 el mismo dato lleva a “rechazar” al 5 % y a “no rechazar” al 1 %. ¿Es eso una contradicción? ¿Qué diría un lector que solo conociera el intervalo al 95 %, y qué le faltaría por saber?
5. En la Actividad 5 el coeficiente sigue siendo significativo con doce viviendas. Un compañero concluye que “entonces con doce viviendas basta”. ¿Qué le respondería, con el intervalo en la mano?

Para profundizar

- Lección 12 ([lección 12](#)), en especial las secciones “Cómo se lee una tabla de resultados” y “Significación estadística $\neq$ relevancia económica”.
- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 4, secciones 4.2 (contraste t), 4.3 (intervalos de confianza) y 4.6 (cómo se informan los resultados).
- Manual de Gretl: funciones `critical` y `pvalue`; comando `restrict`.

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 hprice2.gdt --quiet

outfile --quiet tabla.txt
    ols price const rooms
end outfile

outfile --quiet reconstruccion.txt
    scalar n = $nobs
    scalar k = 2                                # regresores, constante incluida
    scalar SRC = $ess                           # suma de cuadrados de los residuos
    scalar STC = SRC/(1 - $rsq)                 # Gretl no da STC directamente;  $R^2 = 1 - SRC/STC$ 
    scalar SEC = STC - SRC
    scalar s = sqrt(SRC/(n-k))                  # raíz de la cuasivarianza de los residuos
    scalar Lx = sqrt(sum((rooms - mean(rooms))^2)) # longitud euclídea de  $x - \bar{x}$ 
    scalar ee = s/Lx                           # error estándar: ruido por unidad de longitud
    scalar b2 = $coeff(rooms)
    scalar tst = b2/ee
    scalar tc = critical(t, n-k, 0.025)         # valor crítico al 5%, dos colas
    scalar p = 2*pvalue(t, n-k, abs(tst))       # valor p bilateral

    printf "n - k" = %d\n", n-k
    printf "s (D.T. de la regresion)" = %.4f Gretl: %.4f\n", s, $sigma
    printf "||x - xbar||" = %.4f\n", Lx
    printf "ee(b2) = s/||x - xbar||" = %.4f Gretl: %.4f\n", ee, $stderr(rooms)
    printf "t = b2/ee" = %.4f\n", tst
    printf "valor critico t_c (5%%)" = %.4f\n", tc
    printf "valor p" = %.3g\n", p
    printf "IC 95%%: [%.2f ; %.2f]\n", b2 - tc*ee, b2 + tc*ee
end outfile

outfile --quiet identidad.txt
    printf "sqrt((n-k)*SEC/SRC) = %.4f t = %.4f\n", sqrt((n-k)*SEC/SRC), tst
    printf "(n-k)*R2/(1-R2) = %.4f t^2 = %.4f\n", (n-k)*$rsq/(1-$rsq), tst^2
    printf "F(1, %d) del pie de la tabla = %.4f\n", n-k, $Fstat
end outfile

outfile --quiet restriccionCero.txt
    restrict
        b[2] = 0
    end restrict
end outfile
```

```

outfile --quiet contrasteC.txt
    scalar c      = 10000
    scalar t_c0 = (b2 - c)/ee
    scalar p_c0 = 2*pvalue(t, n-k, abs(t_c0))
    scalar tc1 = critical(t, n-k, 0.005)           # valor critico al 1%, dos colas

    printf "H0: beta_2 = %d\n", c
    printf "t = (b2 - c)/ee      = %.4f\n", t_c0
    printf "valor p              = %.4f\n", p_c0
    printf "t_c al 5%%          = %.4f  -> %s\n", tc, (abs(t_c0) > tc ? "se rechaza" : "no se
    ↪ rechaza")
    printf "t_c al 1%%          = %.4f  -> %s\n", tc1, (abs(t_c0) > tc1 ? "se rechaza" : "no se
    ↪ rechaza")
    printf "IC 95%%: [%.2f ; %.2f] ¿contiene a %d? %s\n", b2 - tc*ee, b2 + tc*ee, c, (b2 - tc*ee <= c
    ↪ && c <= b2 + tc*ee ? "si" : "no")
    printf "IC 99%%: [%.2f ; %.2f] ¿contiene a %d? %s\n", b2 - tc1*ee, b2 + tc1*ee, c, (b2 - tc1*ee <= c
    ↪ && c <= b2 + tc1*ee ? "si" : "no")
end outfile

outfile --quiet doceViviendas.txt
    set seed 20261107                                # para que todos obtengamos las mismas doce
    smpl 12 --random
    ols price const rooms
    scalar tc12 = critical(t, $df, 0.025)
    printf "\nn = %d, grados de libertad = %d, t_c al 5%% = %.3f (con 506 viviendas era %.3f)\n", $nobs,
    ↪ $df, tc12, tc
    printf "IC 95%% con 12 viviendas : [%.1f ; %.1f]\n", $coeff(rooms) - tc12*$stderr(rooms), $coeff(rooms)
    ↪ + tc12*$stderr(rooms)
    printf "IC 95%% con 506 viviendas: [%.1f ; %.1f]\n", b2 - tc*ee, b2 + tc*ee
    smpl full
end outfile

```

Enlace al guión: [S18-Prct-A-hprice2-inferencia.inp](#)

Respuestas

1. *Las tres “desviaciones típicas” y el tamaño muestral.* La desviación típica de la variable dependiente ($\sqrt{\text{STC}/(n-1)}$) y la de la regresión ($\mathfrak{s} = \sqrt{\text{SRC}/(n-k)}$) describen la dispersión de los *datos* y del *ruido*: con más viviendas de la misma ciudad, ambas se estabilizarían en torno a un valor, sin tender a cero, porque miden algo que existe en la población (cuánto varían los precios; cuánto ruido hay alrededor de la recta). El error estándar del coeficiente, en cambio, es $\mathfrak{s}/\|\mathbf{x} - \bar{\mathbf{x}}\|_e$, y su denominador *crece* con el número de observaciones (más sumandos en $\sum_i (x_i - \mu_{\mathbf{x}})^2$): con más datos, el error estándar se reduce hacia cero. Esa es la diferencia entre una dispersión de datos y una dispersión de un estimador, y es la razón por la que compartir etiqueta es tan mala idea.
2. *Cambio de unidades en el regresor.* Si $\mathbf{x}$ se mide en decenas de habitaciones, el vector es $\mathbf{x}/10$ y su longitud en desviaciones se divide por diez (homogeneidad de la norma, lección 2). $\mathfrak{s}$ no cambia, porque el ajuste —y por tanto el residuo— es el mismo: solo cambia el nombre de la pendiente, que pasa a ser $10\hat{\beta}_2$. Así que el error estándar se multiplica por diez, igual que el coeficiente, y el cociente t se conserva. Es la pregunta 10 de la lección 12, y la práctica B de hoy la ejecuta con números.
3. *Qué hay en F que no esté en R^2 .* El factor $n - k$, y nada más. El F de una regresión simple es $(n-k) R^2/(1-R^2)$: el mismo cociente de catetos, multiplicado por el número de dimensiones disponibles para el residuo. Por eso un R^2 de 0,48 —modesto— viene acompañado de un F de 473: con quinientas observaciones, un ángulo modesto es una evidencia enorme. La lección 13 explotará el F para juzgar varios regresores a la vez, donde el R^2 por sí solo ya no basta.
4. *Rechazar al 5 % y no al 1 %.* No es una contradicción: son dos exigencias distintas ante la misma evidencia, y el valor $p = 0,036$ las resume a las dos —es más pequeño que 0,05 y más grande que 0,01—. Quien solo tuviera el intervalo al 95 % vería que 10 000 queda fuera y concluiría correctamente “se rechaza al 5 %”; lo que no podría saber es por cuánto: si 10 000 está justo fuera o muy lejos. Para eso necesitaría el valor p , o bien el intervalo al 99 %, que aquí sí lo contiene. El intervalo y el contraste son el mismo objeto, pero un intervalo a un solo nivel es una única rebanada de esa información.
5. *“Con doce viviendas basta”.* Basta para rechazar que el efecto sea cero, que es una hipótesis muy poco ambiciosa: nadie discutía que las casas con más habitaciones valgan más. No basta para nada más: el intervalo, de unos 1 900 a casi 20 000 dólares por habitación, es compatible con que una habitación valga la mitad que en la estimación completa o el doble. Significación no es precisión. Si la pregunta es “¿cuánto vale una habitación?”, doce viviendas responden “entre poco y mucho”, y eso es no responder.