

Índice general

1	De dónde venimos: los contrastes t son individuales	2
2	La hipótesis conjunta	3
3	El estadístico F	4
4	Lectura geométrica: la misma figura	6
5	El F en una tabla de resultados	8
6	Restricciones lineales: modelos anidados	9
7	Tres restricciones sobre <code>hprice2</code>	11
8	Conjunto $\neq$ individual: el caso de la apertura	12
9	Qué dice y qué no dice el F	14
10	Recapitulación y guiño a las sesiones siguientes	15

Lección 13. El contraste F : significación conjunta

Marcos Bujosa

19 de septiembre de 2026

Resumen

Hoy juzgamos varios coeficientes a la vez: ¿aportan algo *en conjunto* estos regresores?, ¿cumplen esta relación? El estadístico F es la razón entre los dos catetos del triángulo de la lección 8, cada uno dividido por la dimensión del subespacio donde vive. La segunda mitad se dedica a leerlo.

“Varios coeficientes, una hipótesis: hace falta un estadístico que mire varias direcciones a la vez.”

- ([slides](#)) — ([html](#)) — ([pdf](#))

1 De dónde venimos: los contrastes t son individuales

La lección 12 juzga *un* coeficiente cada vez, $t = \hat{\beta}_j / \text{ee}(\hat{\beta}_j)$, y demostró (con $k = 2$) que $t^2 = (n - k) \text{SEC}/\text{SRC}$. En el laboratorio anterior, al pie de la tabla de `price` sobre `rooms`, apareció un $F(1, 504) = 472,95$ igual a t^2 .

Un caso real (`hprice2`; `price` sobre `rooms`, `nox` y su logaritmo `lnox`):

Regresor	Coficiente	Desv. típica	Estadístico t	Valor p
<code>rooms</code>	8161,84	418,34	19,51	10^{-63}
<code>nox</code>	404,19	2198,85	0,18	0,854
<code>lnox</code>	-13262,2	12655,7	-1,05	0,295

¿Sobran `nox` y `lnox`? Cada t , por separado, dice que sí. Los dos a la vez, *no*. Hoy: el F por derecho propio, y la resolución de este caso al final.

De dónde venimos: los contrastes t son individuales

Todo el aparato de la lección 12 —error estándar, intervalo de confianza, contraste t — se pronuncia sobre *un* coeficiente cada vez. Es lo natural cuando la pregunta es “¿interviene este regresor?”, y es lo que hace la tabla de resultados fila a fila. Pero muchas preguntas económicas no son sobre un coeficiente, sino sobre varios a la vez: ¿aportan algo *en conjunto* estos tres regresores?, ¿tienen el mismo efecto estas dos variables?, ¿son compatibles los datos con que la suma de dos coeficientes valga uno? Ninguna de ellas se responde con una fila de la tabla, y hoy construimos la herramienta que las responde a todas.

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

Dos piezas de las sesiones anteriores hacen casi todo el trabajo. La primera es la identidad de la lección 12, $t^2 = (n - k) \text{SEC}/\text{SRC}$, que decía que el contraste t es la razón entre los dos catetos del triángulo de la lección 8, corregida por la dimensión. La segunda es una observación del laboratorio anterior: al pie de toda tabla de regresión Gretl imprime un número rotulado F , y en la regresión simple de `price` sobre `rooms` ese número —472,95— coincidía con t^2 . Allí dijimos que el F era “el cociente entre la parte explicada y la parte residual, cada una dividida por su número de dimensiones” y que la lección 13 lo convertiría en un contraste por derecho propio. Esa es la lección de hoy.

La tabla de la transparencia es el caso que da sentido a todo lo demás, y conviene mirarlo despacio antes de saber resolverlo. Es una regresión con datos reales —las viviendas de `hprice2` de los laboratorios— del precio mediano sobre tres regresores: el número de habitaciones y *dos* medidas de la contaminación del aire, la concentración de óxidos de nitrógeno `nox` y su logaritmo `lnox`. Los coeficientes de `nox` y `lnox` tienen ambos un valor p muy alto: 0,854 y 0,295. Si uno lee la tabla como enseñó la lección 12, fila a fila, concluye que ninguno de los dos interviene y que ambos sobran. Y sin embargo, como veremos en la sección “Conjunto $\neq$ individual”, si se quitan *los dos a la vez* el ajuste empeora de forma inconfundible: la contaminación sí explica el precio. Lo que falla no es la tabla; es leer una pregunta sobre dos coeficientes con dos respuestas sobre uno.

Una advertencia sobre el reparto de la sesión, como en la lección 12. Las cuatro primeras transparencias construyen el contraste F de significación conjunta y lo leen en el triángulo; la sexta y la séptima lo generalizan a cualquier hipótesis lineal sobre los coeficientes; las tres últimas se dedican a interpretarlo, y son, otra vez, las de más rendimiento práctico.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, introducción a la sección 4.5 (por qué los contrastes t individuales no responden a hipótesis sobre varios coeficientes).

2 La hipótesis conjunta

Significación conjunta: $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$ (“ningún regresor aporta nada”) frente a H_1 : al menos uno es distinto de cero.

Geométricamente: bajo H_0 el modelo se reduce a la constante, el ajuste es el vector de medias, $\hat{\mathbf{y}} = \bar{\mathbf{y}}$, y el cateto explicado es nulo. La pregunta es si el subespacio de los regresores acerca $\hat{\mathbf{y}}$ a $\mathbf{y}$ más de lo que el azar acercaría.

¿Por qué no $k - 1$ contrastes t ? Cada t juzga a su regresor *con los demás dentro del modelo*; “todos no significativos uno a uno” no es “todos nulos a la vez” (la tabla de la apertura). Y $k - 1$ veredictos al 5 % no forman un veredicto al 5 %.

Una hipótesis sobre varios coeficientes pide un estadístico que mire varias direcciones a la vez.

La hipótesis conjunta

La hipótesis nula más frecuente sobre varios coeficientes es la de *significación conjunta*: que todos los coeficientes de los regresores no constantes sean cero a la vez, $\beta_2 = \dots = \beta_k = 0$. En términos del modelo poblacional de la lección 10, $Y = \beta_1 1 + \beta_2 X_2 + \dots + \beta_k X_k + U$, dice que ninguno de los regresores interviene: $Y = \beta_1 1 + U$, y lo mejor que puede hacerse para predecir Y

es su media. La alternativa es la negación lógica: *al menos uno* de esos coeficientes es distinto de cero. Nótese lo poco que pide H_1 : no dice cuál ni cuántos, solo que no son todos nulos.

La traducción geométrica es inmediata con el vocabulario de la lección 10. El ajuste $\hat{\mathbf{y}}$ es la proyección de $\mathbf{y}$ sobre $\mathcal{L}(\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_k)$, y el triángulo rectángulo de la lección 8 —hipotenusa $\mathbf{y} - \bar{\mathbf{y}}$, catetos $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ y $\hat{\mathbf{e}}$ — sigue existiendo sea cual sea k , como se comprobó allí. Si H_0 fuese cierta, el modelo poblacional no contendría ningún regresor y el ajuste que mejor lo reproduce sería el vector de medias: el cateto explicado $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ “debería” ser nulo. No lo será exactamente, claro: aunque los regresores no tengan nada que ver con $\mathbf{y}$, la proyección de un vector de ruido sobre el subespacio de los regresores en desviaciones nunca es exactamente cero, igual que en la lección 12 la proyección sobre un único regresor no lo era. La pregunta que el contraste tiene que responder es, por tanto, si el cateto explicado es *más largo de lo que el azar produciría* cuando H_0 es cierta. Es la misma pregunta de la lección 12, con una diferencia: allí el cateto explicado vivía en una recta (la del regresor en desviaciones) y aquí vive en un subespacio de dimensión $k - 1$ (el de los $k - 1$ regresores en desviaciones, lección 10).

Conviene detenerse en por qué no basta con mirar los $k - 1$ contrastes t de la tabla, porque es el error de lectura más frecuente y la tabla de la apertura lo exhibe. Hay dos razones distintas. La primera es de *contenido*: cada t contrasta $H_0: \beta_j = 0$ *manteniendo a los demás regresores en el modelo*. Es decir, el t de `nox` pregunta si `nox` aporta algo *además de* `rooms` y `lnox`, y el de `lnox` pregunta si aporta algo *además de* `rooms` y `nox`. Ambos pueden responder “no” —cada uno es prescindible si el otro se queda— sin que los dos sean prescindibles a la vez. “Todos individualmente no significativos” y “todos nulos simultáneamente” son afirmaciones distintas, y la segunda es la que H_0 formula. La segunda razón es de *probabilidad*: cada contraste t al 5 % tiene, bajo su hipótesis nula, un 5 % de probabilidad de rechazar por azar; si se hacen $k - 1$ y se rechaza H_0 conjunta en cuanto alguno rechace, la probabilidad de un rechazo espurio ya no es el 5 %, sino mayor, y tanto mayor cuantos más regresores haya. Un contraste conjunto es un único contraste, con un único nivel, sobre la hipótesis que de verdad interesa.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (“Testing multiple linear restrictions”: el planteamiento de la hipótesis conjunta y por qué los t individuales no la contrastan).

3 El estadístico F

Los dos catetos de la lección 8, cada uno dividido por la dimensión del subespacio donde vive ($k - 1$ para el explicado, $n - k$ para el residuo, lección 11):

$$F = \frac{\text{SEC}/(k - 1)}{\text{SRC}/(n - k)} \quad (\text{el denominador es } \mathfrak{s}^2).$$

Bajo H_0 , y con el ruido isótropo de la lección 12, la longitud al cuadrado del ruido se reparte por igual entre las dimensiones: numerador y denominador estiman los dos σ^2 , y $F \approx 1$. Un F mucho mayor que 1 desmiente H_0 .

Enunciado, no demostrado: bajo H_0 , $F \sim F_{k-1, n-k}$. Se rechaza al nivel α si $F > F_c$, con F_c el valor que deja α en la *cola derecha* (una sola cola: F compara longitudes, no tiene signo). Valor p : $P(F_{k-1, n-k} > F)$.

El estadístico F

La lección 12 dejó el estadístico t escrito como $t^2 = (n - k) \text{SEC} / \text{SRC}$, y en su recapitulación anunció que el cociente SEC / SRC admitía “una versión más general en la que cada una se divide por sus propios grados de libertad”. Esa versión es el estadístico F :

$$F = \frac{\text{SEC} / (k - 1)}{\text{SRC} / (n - k)}.$$

Los dos divisores no son arbitrarios, y el curso ya conoce a los dos. El denominador, $\text{SRC} / (n - k)$, es la cuasivarianza de los residuos $\mathbf{s}^2$ de la lección 11: la suma de cuadrados del residuo dividida por la dimensión $n - k$ del subespacio $\mathcal{L}(\mathbf{X})^\perp$ donde el residuo vive. El numerador hace lo mismo con el otro cateto: $\text{SEC} = \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|_e^2$ es la suma de cuadrados del cateto explicado, y $k - 1$ es la dimensión del subespacio donde ese cateto vive —el de los regresores en desviaciones, que en la lección 10 vimos que tiene dimensión $k - 1$ dentro de $\mathcal{L}(\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_k)$ —. Así que F compara *longitud al cuadrado por dimensión* en los dos catetos: cuánta longitud al cuadrado “le toca” a cada una de las $k - 1$ direcciones del subespacio explicado, frente a cuánta le toca a cada una de las $n - k$ direcciones del residuo.¹

¿Por qué es ese cociente, y no otro, el que sirve para contrastar H_0 ? La respuesta es la misma idea de isotropía que en la lección 12 explicaba por qué el valor crítico depende de $n - k$, y conviene enunciarla con cuidado, porque no vamos a demostrarla. Supongamos H_0 cierta. Entonces el modelo poblacional es $Y = \beta_1 \mathbf{1} + U$: el vector de datos en desviaciones, $\mathbf{y} - \bar{\mathbf{y}}$, no es más que el vector de perturbaciones centrado, y bajo los supuestos de la lección 12 (homocedasticidad, independencia, normalidad) ese vector es *isótropo* dentro de $\mathcal{L}(\mathbf{1})^\perp$: no privilegia ninguna dirección. Un vector isótropo con varianza σ^2 por componente reparte su longitud al cuadrado *por igual* entre las dimensiones. Proyectado sobre cualquier subespacio de dimensión d , la longitud al cuadrado de su proyección vale, en promedio, $d\sigma^2$.² Aplicado a los dos catetos: bajo H_0 , la longitud al cuadrado del cateto explicado vale en promedio $(k - 1)\sigma^2$ y la del residuo, $(n - k)\sigma^2$.³ Dividiendo cada una por su dimensión, *numerador y denominador de F estiman ambos σ^2* , y su cociente debe rondar 1. Si, en cambio, H_0 es falsa, el cateto explicado contiene además la parte de $\mathbf{y}$ que los regresores sí explican —una longitud que no es ruido y que no se reparte con el residuo—, el numerador crece y F se hace grande. Un F mucho mayor que 1 es, por tanto, evidencia contra H_0 ; un F cercano a 1 (o menor) es compatible con ella.

Para convertir esta idea en un contraste con nivel de significación hace falta saber *cuánto* mayor que 1 es “mucho”, y eso exige la distribución de F bajo H_0 . Igual que en la lección 12 con la t

¹Con este lenguaje, los tres divisores de la familia de productos escalares del curso —1, n y $n - k$ — se ven acompañados hoy por un cuarto número, $k - 1$, que no da lugar a ningún producto escalar nuevo: es sencillamente la dimensión del otro cateto. Todas las dimensiones que aparecen en esta lección suman lo que deben: $(k - 1) + (n - k) = n - 1$, la dimensión de $\mathcal{L}(\mathbf{1})^\perp$ (lección 5), el plano de las desviaciones donde vive el triángulo entero.

²Es un resultado que puede demostrarse con las herramientas de la lección 11 y no lo haremos; su versión más elemental ya la conoce: la longitud al cuadrado de un vector de n componentes independientes con varianza σ^2 y media cero vale $n\sigma^2$ en promedio, porque cada componente aporta σ^2 . Proyectar sobre un subespacio de dimensión d deja “ d componentes” de las n , y la isotropía garantiza que da igual cuáles.

³Las palabras “en promedio” esconden el mismo abuso de notación de la lección 11: SEC y SRC son aquí las fórmulas de la lección 8 *evaluadas sobre la muestra aleatoria* —variables aleatorias, no los números de una tabla—, y lo que se afirma es $E(\text{SEC}) = (k - 1)\sigma^2$ bajo H_0 y $E(\text{SRC}) = (n - k)\sigma^2$. La segunda igualdad es, dicho de otro modo, la insesgadez de $\mathbf{s}^2$ que la lección 11 enunció; y vale con o sin H_0 , porque el residuo no cambia por que los regresores aporten algo o no.

de Student, la enunciamos sin demostrar: bajo H_0 y los supuestos de la lección 12, el estadístico sigue una *distribución F de Fisher–Snedecor con $k - 1$ y $n - k$ grados de libertad*, $F_{k-1, n-k}$. Es una distribución que solo toma valores positivos, asimétrica, con la masa concentrada en torno a 1 y una cola derecha larga. Su forma depende de los dos números de grados de libertad, que son, una vez más, las dos dimensiones del triángulo.⁴ La regla de decisión tiene una novedad respecto a la lección 12: la región crítica está en *una sola cola*. Se rechaza H_0 al nivel α si $F > F_c$, donde F_c es el valor de la $F_{k-1, n-k}$ que deja probabilidad α a su derecha. No hay cola izquierda porque F compara longitudes al cuadrado y no tiene signo: un F pequeño dice que el cateto explicado es corto, que es precisamente lo que H_0 predice, y nunca puede ser evidencia contra ella. El valor p es, como siempre, la probabilidad de un valor al menos tan grande como el observado si H_0 fuese cierta, $P(F_{k-1, n-k} > F)$, y Gretl lo imprime al lado del F al pie de cada tabla.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (el estadístico F y su distribución bajo la hipótesis nula); apéndice B (distribuciones χ^2 , t y F y la relación $t_\nu^2 = F_{1, \nu}$).

4 Lectura geométrica: la misma figura

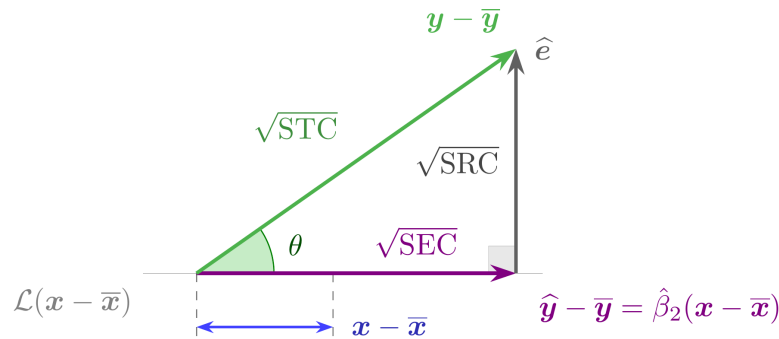


Figura 1: El triángulo de las lecciones 8 y 12, sin cambiar un trazo. En verde, el vector de datos en desviaciones (hipotenusa, longitud $\sqrt{\text{STC}}$); en violeta, el ajuste en desviaciones (cateto explicado, longitud $\sqrt{\text{SEC}}$), que con varios regresores ya no es múltiplo de ninguno de ellos y vive en un subespacio de dimensión $k - 1$; en gris, el residuo (cateto residual, longitud $\sqrt{\text{SRC}}$), en un subespacio de dimensión $n - k$. El ángulo θ es el de $R^2 = \cos^2 \theta$. El estadístico F compara los dos catetos, cada uno dividido por su dimensión.

$$F = \frac{n - k}{k - 1} \cdot \frac{\text{SEC}}{\text{SRC}} = \frac{n - k}{k - 1} \cdot \frac{R^2}{1 - R^2} : \quad \text{función únicamente del ángulo } \theta \text{ y de dos dimensiones.}$$

⁴Su prueba exige la distribución χ^2 de cada suma de cuadrados y la independencia entre ambas, el mismo aparato que la lección 12 declinó desarrollar para la t . De hecho, la t_{n-k} y la $F_{1, n-k}$ son la misma distribución mirada de dos maneras: si $T \sim t_{n-k}$, entonces $T^2 \sim F_{1, n-k}$. Es la versión distribucional de $F = t^2$ que veremos en la transparencia siguiente.

Con $k = 2$: $F = (n - 2) \text{SEC}/\text{SRC} = t^2$, y también $F_c = t_c^2$. El ejemplo de la lección 8 ($n = 5$): $F = \frac{90/1}{10/3} = 27 = t^2$; el valor crítico $F_c(1, 3) = 10,13 = 3,18^2$.

Lectura geométrica: la misma figura

La figura es la de la lección 12, y no es una economía de medios: es que el objeto es el mismo. Con k regresores el triángulo rectángulo de la lección 8 sigue existiendo —la lección 10 lo comprobó: $\hat{\mathbf{e}} \perp \hat{\mathbf{y}} - \bar{\mathbf{y}}$ porque el residuo es ortogonal a todo el subespacio de los regresores y ese vector está en él—, y $\text{STC} = \text{SEC} + \text{SRC}$ y $R^2 = \cos^2 \theta$ valen sin cambio. Lo único que cambia respecto a la lección 12 es una frase del pie de figura: el cateto explicado ya no es un múltiplo del regresor, porque ahora hay $k - 1$ regresores en desviaciones y $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ es una combinación de todos ellos que, en general, no está alineada con ninguno (lección 10). Por eso la figura de la lección 12 medía el regresor bajo el eje y esta no: la recta horizontal representa aquí un subespacio de dimensión $k - 1$, no una dirección.

Escribamos F en términos del triángulo. Sacando los divisores fuera,

$$F = \frac{\text{SEC}/(k - 1)}{\text{SRC}/(n - k)} = \frac{n - k}{k - 1} \cdot \frac{\text{SEC}}{\text{SRC}},$$

y como $\text{SEC}/\text{SRC} = R^2/(1 - R^2)$ (lección 12, dividiendo numerador y denominador por STC y usando Pitágoras),

$$F = \frac{n - k}{k - 1} \cdot \frac{R^2}{1 - R^2}.$$

Es la misma estructura que la identidad de la lección 12 con un factor distinto: allí $t^2 = (n - k) R^2/(1 - R^2)$; aquí el $n - k$ se divide además por $k - 1$. En los dos casos el estadístico es una función únicamente del ángulo θ —vía $R^2 = \cos^2 \theta$ — y de dimensiones. Y en los dos casos vale la misma lectura: *el contraste no mide nada que el R^2 no midiera ya*; lo que añade es el factor que cuenta cuántas dimensiones respaldan ese ángulo. Un R^2 dado es tanto más sorprendente cuantas más observaciones haya ($n - k$ grande) y cuantos *menos* regresores se hayan usado para conseguirlo ($k - 1$ pequeño). Esta segunda parte es nueva respecto a la lección 12, y es la intuición que la práctica sobre el R^2 ajustado ya manejó: con más regresores es más fácil acercar $\hat{\mathbf{y}}$ a $\mathbf{y}$ por puro azar, porque el subespacio sobre el que se proyecta es mayor, y el F lo descuenta dividiendo por $k - 1$.

El caso $k = 2$ cierra el círculo con la lección 12. Con un único regresor no constante, $k - 1 = 1$ y la fórmula se reduce a $F = (n - 2) \text{SEC}/\text{SRC}$, que es t^2 . No solo el estadístico: también el valor crítico, porque la $F_{1, n-2}$ es la distribución del cuadrado de una t_{n-2} , de modo que $F_c = t_c^2$ y las dos reglas de decisión rechazan en los mismos casos. Es lo que el laboratorio anterior verificó sobre `hprice2` y con `restrict`, y lo que el ejemplo de juguete de la lección 8 permite comprobar con números redondos: $\text{SEC} = 90$, $\text{SRC} = 10$, $n = 5$, luego $F = (90/1)/(10/3) = 27$, el mismo $t^2 = 27$ de la lección 12; y el valor crítico al 5% de la $F_{1,3}$ es 10,13, el cuadrado del 3,18 de la t_3 . Se rechaza por las dos vías, y por el mismo margen.

También la región crítica se lee igual que en la lección 12: rechazar cuando $F > F_c$ equivale a rechazar cuando el ángulo θ es menor que un ángulo crítico θ_c —un *cono* alrededor del subespacio de los regresores en desviaciones—, y la abertura de ese cono depende de las dos dimensiones. Con

los números de la transparencia siguiente ($n = 506$, $k = 4$), la abertura al 5 % es de $82,9^\circ$: casi todo el espacio, por la misma razón que en la lección 12 —en dimensión alta el azar solo produce ángulos próximos a 90° —.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5, apartado “The R -squared form of the F statistic”.

5 El F en una tabla de resultados

hprice2: price sobre rooms, nox y stratio (el modelo de la práctica que siguió a la lección 10).

Regresor	Coficiente	Desv. típica	Estadístico t
constante	9458,01	4371,41	2,16
rooms	7000,63	409,15	17,11
nox	-1693,33	236,25	-7,17
stratio	-1166,96	128,76	-9,06

$n = 506$, $k = 4$, $R^2 = 0,6005$, $\text{SRC} = 1,711 \cdot 10^{10}$, $\text{SEC} = 2,572 \cdot 10^{10}$, $\mathfrak{s} = 5837,8$; al pie: $F(3, 502) = 251,5$, valor $p = 1,3 \cdot 10^{-99}$.

Reconstrucción: $\text{SEC}/3 = 8,573 \cdot 10^9$ y $\text{SRC}/502 = 3,408 \cdot 10^7 = \mathfrak{s}^2$; cociente 251,5, frente a $F_c(3, 502) = 2,62$ al 5 %. *Dice:* los tres coeficientes no son cero a la vez. *No dice:* cuál importa, ni cuánto.

El F en una tabla de resultados

Toda tabla de regresión de Gretl termina con dos líneas de resumen que, hasta hoy, habíamos leído solo en parte: el R^2 y la desviación típica de la regresión $\mathfrak{s}$ (lección 12), y a su lado un “ $F(k-1, n-k)$ ” con su valor p . Ese F es el estadístico de significación conjunta de esta lección: contrasta, automáticamente y para cualquier modelo que se estime, la hipótesis de que todos los coeficientes de los regresores no constantes son cero a la vez. Los dos números entre paréntesis son las dimensiones del triángulo.

Tomemos el modelo de la práctica que siguió a la lección 10, el precio mediano de la vivienda sobre el número de habitaciones, la contaminación y la ratio alumnos/profesor del distrito. Con $n = 506$ y $k = 4$ (tres regresores más la constante), las dimensiones son $k-1 = 3$ y $n-k = 502$. Todo lo que hace falta para reconstruir el F está en la tabla o sale de ella con la lección 8: SRC lo imprime Gretl; STC se obtiene de $R^2 = 1 - \text{SRC}/\text{STC}$, y $\text{SEC} = \text{STC} - \text{SRC} = 2,572 \cdot 10^{10}$. Dividiendo cada cateto por su dimensión, el numerador vale $8,573 \cdot 10^9$ y el denominador $3,408 \cdot 10^7$, que es $\mathfrak{s}^2 = 5837,8^2$, como debe. El cociente es 251,5, el F de la tabla. Por la otra vía, $\frac{502}{3} \cdot \frac{0,6005}{0,3995} = 251,5$, el mismo número.

La decisión no admite duda: el valor crítico de la $F_{3,502}$ al 5 % es 2,62, y al 1 %, 3,82; el estadístico observado es cien veces mayor y el valor p es del orden de 10^{-99} . Se rechaza H_0 : los datos no son compatibles con que **rooms**, **nox** y **stratio** tengan, a la vez, coeficiente nulo. Merece la pena leer también la magnitud: $F = 251,5$ dice que, por dimensión, el cateto explicado tiene una longitud al

cuadrado doscientas cincuenta veces mayor que la del residuo; si H_0 fuese cierta esa razón debería rondar 1.

Y ahora lo que el F *no* dice, que es tan importante como lo que dice y que la sección “Qué dice y qué no dice el F ” desarrollará. Rechazar $H_0: \beta_2 = \beta_3 = \beta_4 = 0$ es rechazar que *los tres* sean nulos a la vez; es compatible con que dos de ellos lo sean. El F conjunto no señala a ningún regresor en particular, ni dice nada de la magnitud de los efectos: para eso están las columnas de la tabla y los intervalos de confianza de la lección 12. En este ejemplo los tres t individuales son grandes y no hay conflicto; pero la tabla de la apertura muestra que puede haberlo, y en ese caso la lectura correcta exige entender qué mide cada estadístico.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5, apartado “The F statistic for overall significance of a regression”; y sección 4.6 (cómo se presentan los resultados, incluido el F).

6 Restricciones lineales: modelos anidados

Una hipótesis lineal: q ecuaciones sobre los β (excluir, igualar, fijar). Imponerla es proyectar sobre un subespacio *menor*: ajuste restringido $\hat{\mathbf{y}}_r$, con $\text{SRC}_r \geq \text{SRC}$.

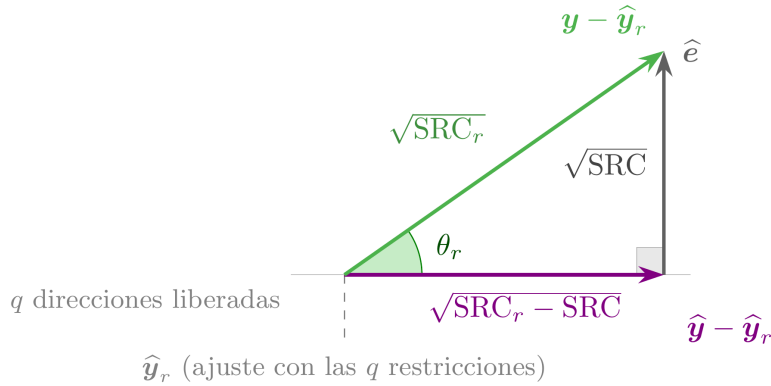


Figura 2: El mismo triángulo con el ajuste restringido en el papel del vector de medias. En verde, el residuo del modelo restringido $\mathbf{y} - \hat{\mathbf{y}}_r$ (hipotenusa, longitud $\sqrt{\text{SRC}_r}$); en violeta, lo que las q direcciones liberadas añaden al ajuste, $\hat{\mathbf{y}} - \hat{\mathbf{y}}_r$ (cateto de longitud $\sqrt{\text{SRC}_r - \text{SRC}}$); en gris, el residuo del modelo completo (cateto de longitud $\sqrt{\text{SRC}}$). El ángulo recto es la ortogonalidad $\hat{\mathbf{y}} - \hat{\mathbf{y}}_r \perp \hat{\mathbf{e}}$, de la que sale Pitágoras: $\text{SRC}_r = \text{SRC} + \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_r\|_{\hat{\mathbf{e}}}^2$. Con $\hat{\mathbf{y}}_r = \bar{\mathbf{y}}$ es la figura anterior.

$$F = \frac{(\text{SRC}_r - \text{SRC})/q}{\text{SRC}/(n-k)} \sim F_{q, n-k} \text{ bajo } H_0 \quad (\text{conjunta: } \hat{\mathbf{y}}_r = \bar{\mathbf{y}}, q = k-1).$$

Restricciones lineales: modelos anidados

La significación conjunta es solo una de las hipótesis que pueden formularse sobre varios coeficientes, y no siempre la más interesante. Una *hipótesis lineal* general es un sistema de q ecuaciones lineales sobre los coeficientes del modelo poblacional. Tres ejemplos que reaparecerán con números en la transparencia siguiente: excluir dos regresores a la vez, $\beta_3 = \beta_4 = 0$ ($q = 2$); excluir uno solo, $\beta_4 = 0$ ($q = 1$, el contraste t de siempre); y una hipótesis que no excluye a nadie, sino que relaciona dos coeficientes, $\beta_3 = \beta_4$ ($q = 1$). Otras, igual de naturales en economía, fijan un valor —una elasticidad unitaria, $\beta_2 = 1$ — o una suma —rendimientos constantes a escala, $\beta_2 + \beta_3 = 1$, que veremos al hablar de la función de producción—.

Lo que hace tratables a todas ellas con una única herramienta es una observación geométrica. Imponer q restricciones lineales sobre los coeficientes equivale a exigir que el ajuste sea una combinación lineal de los regresores *de un tipo particular*. El conjunto de esas combinaciones es un subespacio *menor* que $\mathcal{L}(\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_k)$, contenido en él, con q dimensiones menos.⁵ El *ajuste restringido* $\hat{\mathbf{y}}_r$ es la proyección de $\mathbf{y}$ sobre ese subespacio menor —lo que Gretl calcula al estimar, por ejemplo, el modelo sin `nox` ni `stratio` a la vez—, y su suma de cuadrados residual, $\text{SRC}_r = \|\mathbf{y} - \hat{\mathbf{y}}_r\|_e^2$, es necesariamente mayor o igual que la del modelo completo: proyectar sobre un subespacio más pequeño nunca acerca más. Es la misma propiedad mecánica que la lección 8 anunció como “ R^2 nunca empeora al añadir regresores” y que la práctica del R^2 ajustado hizo explícita: quitar regresores nunca reduce SRC.

La pregunta del contraste es, entonces, *cuánto* ha subido SRC al imponer las restricciones. Si H_0 es cierta, el modelo restringido es el verdadero y la subida $\text{SRC}_r - \text{SRC}$ es solo la parte del ruido que las q direcciones liberadas recogían por azar. Si H_0 es falsa, las restricciones le quitan al ajuste algo que sí explicaba $\mathbf{y}$, y la subida es grande. Para medirla hace falta el único resultado que hoy sí demostramos:

Proposición (Pitágoras anidado). Los dos ajustes cumplen $\hat{\mathbf{y}} - \hat{\mathbf{y}}_r \perp \hat{\mathbf{e}}$, y por tanto

$$\text{SRC}_r = \text{SRC} + \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_r\|_e^2.$$

Demostración. El ajuste restringido $\hat{\mathbf{y}}_r$ es una combinación lineal de los regresores (una de un tipo particular, pero combinación al fin), luego pertenece a $\mathcal{L}(\mathbf{1}, \mathbf{x}_2, \dots, \mathbf{x}_k)$; y $\hat{\mathbf{y}}$ también. Su diferencia está, por tanto, en ese subespacio, y el residuo del modelo completo es ortogonal a todo él (lección 10: $\hat{\mathbf{e}}\mathbf{X} = \mathbf{0}$). Así que $\hat{\mathbf{y}} - \hat{\mathbf{y}}_r \perp \hat{\mathbf{e}}$. Ahora escribimos el residuo restringido como suma de dos vectores ortogonales:

$$\mathbf{y} - \hat{\mathbf{y}}_r = (\hat{\mathbf{y}} - \hat{\mathbf{y}}_r) + (\mathbf{y} - \hat{\mathbf{y}}) = (\hat{\mathbf{y}} - \hat{\mathbf{y}}_r) + \hat{\mathbf{e}}.$$

Aplicando el teorema de Pitágoras de la lección 3, $\|\mathbf{y} - \hat{\mathbf{y}}_r\|_e^2 = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_r\|_e^2 + \|\hat{\mathbf{e}}\|_e^2$, que es la igualdad anunciada. ■

Esta proposición es la de la lección 8 ($\text{STC} = \text{SEC} + \text{SRC}$) con el vector de medias sustituido por el ajuste restringido, y la figura de la transparencia es la de la lección 8 con esa sustitución:

⁵Con $\beta_3 = \beta_4 = 0$, el subespacio restringido es $\mathcal{L}(\mathbf{1}, \mathbf{x}_2)$: hemos perdido dos generadores. Con $\beta_3 = \beta_4$, es $\mathcal{L}(\mathbf{1}, \mathbf{x}_2, \mathbf{x}_3 + \mathbf{x}_4)$: la combinación $\beta_3\mathbf{x}_3 + \beta_4\mathbf{x}_4$ con $\beta_3 = \beta_4$ es $\beta_3(\mathbf{x}_3 + \mathbf{x}_4)$, así que los dos regresores se funden en uno y se pierde una dimensión. Las restricciones que fijan un valor, como $\beta_2 = 1$, dan un conjunto que no pasa por el origen (un subespacio *trasladado* por $\mathbf{x}_2$); el argumento que sigue vale igual, porque lo único que usa es que los dos ajustes están en $\mathcal{L}(\mathbf{X})$.

hipotenusa $\mathbf{y} - \hat{\mathbf{y}}_r$, catetos $\hat{\mathbf{y}} - \hat{\mathbf{y}}_r$ y $\hat{\mathbf{e}}$. La subida de la suma de cuadrados residual es la longitud al cuadrado del cateto que las q direcciones liberadas añaden al ajuste, y ese cateto vive en un subespacio de dimensión q . Con esto, el estadístico se construye igual que el de significación conjunta —cada cateto dividido por su dimensión—:

$$F = \frac{(\text{SRC}_r - \text{SRC})/q}{\text{SRC}/(n - k)} = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_r\|_e^2/q}{\mathbf{s}^2},$$

La misma isotropía de la sección “El estadístico F ” hace que, bajo H_0 , numerador y denominador estimen ambos σ^2 . Su distribución bajo H_0 , que enunciamos sin demostrar, es la $F_{q, n-k}$, y la regla es la de siempre: rechazar si $F > F_c$. La significación conjunta es el caso particular en que el subespacio restringido es $\mathcal{L}(\mathbf{1})$: entonces $\hat{\mathbf{y}}_r = \bar{\mathbf{y}}$, $\text{SRC}_r = \text{STC}$, la subida es SEC y $q = k - 1$. Todo lo dicho en las transparencias anteriores era, sin saberlo, este caso.

Es útil retener el nombre que suele darse a esta construcción: los dos modelos están *anidados*, el restringido dentro del completo, y el F compara el ajuste del pequeño con el del grande. Cualquier par de modelos anidados —cualquier hipótesis lineal— se contrasta con la misma receta: estimar los dos, leer las dos sumas de cuadrados residuales, contar q y $n - k$.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (“Testing exclusion restrictions” y “Testing general linear restrictions”); y sección 4.4 (contraste de una única combinación lineal de coeficientes, el caso $q = 1$).

7 Tres restricciones sobre hprice2

Modelo completo: **price** sobre **rooms**, **nox** y **stratio** ($\text{SRC} = 1,711 \cdot 10^{10}$, $n - k = 502$, $\mathbf{s}^2 = 3,408 \cdot 10^7$). En Gretl: menú **Contrastes -> Restricciones lineales** o el comando **restrict** del laboratorio anterior.

H_0	q	$\text{SRC}_r - \text{SRC}$	F	F_c al 5 %	Veredicto
$\beta_{\text{nox}} = \beta_{\text{stratio}} = 0$	2	$4,985 \cdot 10^9$	73,1	3,01	se rechaza ($p = 10^{-28}$)
$\beta_{\text{stratio}} = 0$	1	$2,799 \cdot 10^9$	82,1	3,86	se rechaza; $82,1 = (-9,06)^2$
$\beta_{\text{nox}} = \beta_{\text{stratio}}$	1	$1,213 \cdot 10^8$	3,56	3,86	no se rechaza ($p = 0,060$)

La segunda fila es el contraste t de la tabla, al cuadrado. La tercera *no* es de exclusión: el modelo restringido regresa **price** sobre **rooms** y la *suma nox + stratio* —otro subespacio, no menos regresores—.

Tres restricciones sobre hprice2

Los tres contrastes de la tabla se hacen sobre el mismo modelo completo de la sección “El F en una tabla de resultados”, de modo que el denominador es siempre el mismo: $\text{SRC}/(n - k) = \mathbf{s}^2 = 3,408 \cdot 10^7$. Lo que cambia de fila a fila es el numerador, es decir, el modelo restringido que se compara.

Primera fila: dos exclusiones a la vez, $H_0 : \beta_{\text{nox}} = \beta_{\text{stratio}} = 0$. El modelo restringido es la regresión simple de **price** sobre **rooms** que arrastramos desde el laboratorio que siguió a la lección 8, con $\text{SRC}_r = 2,209 \cdot 10^{10}$. La subida al quitar los dos regresores es $4,985 \cdot 10^9$, que repartida entre $q = 2$ dimensiones y dividida por $\mathbf{s}^2$ da $F = 73,1$, frente a un valor crítico $F_c(2, 502) = 3,01$. Se

rechaza sin discusión: la contaminación y la ratio alumnos/profesor añaden, juntas, mucho más de lo que dos direcciones de puro azar añadirían. Es la respuesta a la pregunta “¿sobran estas dos variables?”, que ningún t individual formula.

Segunda fila: una exclusión, $H_0: \beta_{\text{stratio}} = 0$. Ahora el modelo restringido es **price** sobre **rooms** y **nox**, con $\text{SRC}_r = 1,991 \cdot 10^{10}$; la subida es $2,799 \cdot 10^9$, $q = 1$, y $F = 82,1$. Compárese con la tabla de la sección “El F en una tabla de resultados”: el t de **stratio** era $-9,06$, y $(-9,06)^2 = 82,1$. Es la identidad $F = t^2$ con $q = 1$, ahora en un modelo con varios regresores: el contraste t de un coeficiente es el contraste F de una única restricción de exclusión, y el valor crítico $F_c(1, 502) = 3,86$ es el cuadrado del $t_c = 1,96$ de la lección 12. Gretl lo confirma si se le pide la restricción con **restrict** o desde el menú: devuelve un $F(1, 502)$ idéntico al t^2 , con el mismo valor p .

Tercera fila: una igualdad, $H_0: \beta_{\text{nox}} = \beta_{\text{stratio}}$ (“una unidad más de contaminación cuesta lo mismo que un alumno más por profesor”). Esta hipótesis no excluye a nadie, y por eso conviene detenerse en cuál es su modelo restringido. Bajo H_0 , la combinación $\beta_{\text{nox}}\mathbf{x}_{\text{nox}} + \beta_{\text{stratio}}\mathbf{x}_{\text{stratio}}$ se convierte en $\beta_{\text{nox}}(\mathbf{x}_{\text{nox}} + \mathbf{x}_{\text{stratio}})$: los dos regresores se funden en uno, la suma, y el modelo restringido es la regresión de **price** sobre **rooms** y **nox + stratio**. Su SRC_r es $1,723 \cdot 10^{10}$, apenas por encima del completo: la subida es $1,213 \cdot 10^8$ y $F = 3,56$, ligeramente por debajo del valor crítico 3,86. No se rechaza al 5 %, con un valor p de 0,060 que conviene informar tal cual: los datos son *compatibles* con que los dos efectos sean iguales, aunque por poco, y la lección 12 ya advirtió que “no rechazar” no es “aceptar”. Lo instructivo de esta fila es que muestra que imponer una restricción es cambiar de subespacio, no necesariamente recortar regresores: el subespacio $\mathcal{L}(\mathbf{1}, \mathbf{x}_{\text{rooms}}, \mathbf{x}_{\text{nox}} + \mathbf{x}_{\text{stratio}})$ tiene una dimensión menos que el completo, pero no se obtiene quitando ninguna columna de $\mathbf{X}$.⁶

En Gretl, las tres filas se obtienen sin estimar a mano el modelo restringido: en la ventana del modelo completo, el menú **Contrastes -> Restricciones lineales** (o el bloque **restrict** que el laboratorio anterior usó con **b[2]=0**) acepta cualquier sistema de ecuaciones lineales sobre los coeficientes y devuelve el F con sus dos grados de libertad y su valor p . Un laboratorio posterior lo aplicará con varias restricciones a la vez y comprobará, estimando los dos modelos, que el F que imprime es el cociente de sumas de cuadrados de esta lección.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.4 (contrastar $\beta_j = \beta_h$ como un t con la covarianza de los estimadores, y por qué la vía del modelo restringido es más cómoda); manual de Gretl, comando **restrict**.

8 Conjunto $\neq$ individual: el caso de la apertura

Modelo (price sobre rooms y ...)	t de nox	t de lnox	R^2	F de excluir ambos
... nox	-7,43		0,5352	
... lnox		-7,51	0,5361	
... nox y lnox	0,18	-1,05	0,5362	$F(2, 502) = 28,2$ ($F_c = 3,01$)

$\rho_{\text{nox}, \text{lnox}} = 0,994$: los dos regresores apuntan casi en la misma dirección. Cada uno solo es muy

⁶Con la fórmula matricial de la lección 11 para la varianza de $\hat{\beta}$, este contraste puede hacerse también como un t : el estadístico es $(\hat{\beta}_{\text{nox}} - \hat{\beta}_{\text{stratio}})$ dividido por el error estándar de esa diferencia, que involucra la covarianza $\text{Cov}[\hat{\beta}_{\text{nox}}, \hat{\beta}_{\text{stratio}} | \mathbf{X}]$ de fuera de la diagonal. Su cuadrado es el F de la tabla; la vía de las sumas de cuadrados evita tener que calcular esa covarianza.

significativo; juntos, el ajuste no puede repartir el efecto entre ellos: a cada uno le queda muy poca longitud *propia* (lección 12), su error estándar se dispara y su t se hunde. Pero el subespacio que generan *juntos* sí acerca $\hat{\mathbf{y}}$ a $\mathbf{y}$, y el F lo ve.

Un F conjunto alto con t individuales nulos es la firma de la *colinealidad* (lección 14). La conclusión correcta no es “sobran los dos”, sino “no se puede separar cuál de los dos”.

Conjunto $\neq$ individual: el caso de la apertura

Volvamos a la tabla con la que abrimos la sesión, ahora con la herramienta para leerla. Las tres filas de la transparencia son tres regresiones del precio mediano de la vivienda: sobre **rooms** y la concentración de óxidos de nitrógeno **nox**; sobre **rooms** y el logaritmo de esa concentración, **lnox**; y sobre **rooms** y las dos medidas a la vez. Las dos primeras dicen lo mismo con claridad: la contaminación reduce el precio, con un t de $-7,4$ o $-7,5$ y un R^2 de $0,535$ o $0,536$. Da casi igual medirla en niveles o en logaritmos, y es fácil ver por qué: la correlación entre **nox** y **lnox** es $0,994$; en el lenguaje de la lección 5, sus vectores en desviaciones forman un ángulo de unos 6° , prácticamente la misma dirección.

La tercera fila es la que desconcierta si se lee fila a fila. Con las dos medidas en el modelo, el t de **nox** cae a $0,18$ y el de **lnox** a $-1,05$; los valores p son $0,854$ y $0,295$, y la lectura ingenua de la lección 12 concluiría que ninguna de las dos interviene. Pero el R^2 apenas ha cambiado $-0,5362-$ y, sobre todo, el F de excluir las dos a la vez vale $28,2$ frente a un crítico de $3,01$: quitar la contaminación del modelo, en cualquiera de sus dos formas, empeora el ajuste de manera inconfundible. Las dos variables *juntas* son claramente significativas; *cada una*, en presencia de la otra, no lo es.

No hay contradicción, y la lección 12 ya dio la clave en su recapitulación: con varios regresores, el error estándar de $\hat{\beta}_j$ es $s/\|\tilde{\mathbf{x}}_j\|_e$, donde $\tilde{\mathbf{x}}_j$ es lo que le queda de *propio* al regresor j tras descontar lo que comparte con los demás. Cuando **nox** y **lnox** están los dos en el modelo, casi todo lo que tiene cada uno lo comparte con el otro. La parte propia de cada uno es un vector cortísimo, el error estándar de cada coeficiente se dispara y el t se hunde: compárense las desviaciones típicas de la tabla de la apertura, $2198,85$ y $12655,7$, con los 236 que tenía **nox** sola en la sección “El F en una tabla de resultados”. El contraste t de **nox** pregunta si **nox** aporta algo *que lnox no aporte ya*, y la respuesta honesta es que casi nada; y lo mismo al revés. Pero esa es una pregunta sobre la *dirección propia* de cada regresor, no sobre el subespacio que generan entre los dos. Ese subespacio —un plano dentro de $\mathcal{L}(\mathbf{1})^\perp$ — sí contiene una dirección en la que $\mathbf{y}$ se proyecta con fuerza, y el F de excluir ambas mide eso: cuánto se alarga el residuo si se suprime el plano entero. Por eso el F es enorme mientras los dos t son nulos.

Esta situación tiene nombre, *colinealidad*, y le dedicaremos la lección 14. Lo que conviene retener hoy es la lectura correcta de la tabla, porque el error contrario es muy frecuente en informes y artículos: ante dos regresores con t pequeños, eliminar los dos “porque no son significativos”. El contraste conjunto dice que eso es un error: lo que los datos no permiten es *separar* el efecto de uno del efecto del otro, no negar que haya efecto. La respuesta razonable es quedarse con uno de los dos —cualquiera de las dos primeras filas es un modelo excelente— o, si la pregunta económica exige distinguirlos, reconocer que estos datos no lo permiten. Es la misma moraleja de la lección 12 con un solo regresor —para medir un efecto hace falta que la variable varíe— trasladada a varios: para medir el efecto de cada uno hace falta que cada uno varíe *por su cuenta*.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5, apartado “Relationship between F and t statistics” (el ejemplo en que ningún t es significativo y el F sí lo es); y cap. 3, sección 3.4, apartado sobre multicolinealidad.

9 Qué dice y qué no dice el F

$F = \frac{n-k}{k-1} \cdot \frac{R^2}{1-R^2}$ mezcla, como el t , fuerza de la relación (el ángulo) con cantidad de evidencia (las dimensiones). Con $k = 4$:

n	R^2	θ	F	Veredicto al 5 %
6	0,90	18,4°	6,00	no significativo ($F_c = 19,16$)
10	0,90	18,4°	18,00	significativo ($F_c = 4,76$)
1003	0,01	84,3°	3,36	significativo ($F_c = 2,61$)

El F del pie de la tabla contrasta la hipótesis *menos* interesante (que nada explique nada) y con muestras grandes se rechaza casi siempre. Los F que informan nacen de una pregunta: ¿sobran estas dos variables?, ¿tienen el mismo efecto?, ¿es la elasticidad unitaria?

Doce viviendas: $F(3, 8) = 9,55 > 4,07$, significativo... con **stratio** a $t = -0,10$ e intervalos enormes. Un F no dice cuál regresor ni cuánto: mire los coeficientes y sus intervalos. Y nada de esto establece causalidad.

Qué dice y qué no dice el F

Todo lo que la lección 12 dijo sobre la lectura del t vale para el F , y la forma $F = \frac{n-k}{k-1} \cdot \frac{R^2}{1-R^2}$ lo hace evidente sin necesidad de moraleja: el estadístico es el producto de un factor que mide la fuerza de la relación en la muestra —el ángulo, el R^2 — y otro que solo cuenta dimensiones. La tabla de la transparencia repite, con $k = 4$, el experimento de la lección 12. Las dos primeras filas tienen el mismo R^2 de 0,90, el mismo ángulo de 18,4°, y veredictos opuestos: con $n = 6$ hay solo $n - k = 2$ dimensiones para el residuo, el valor crítico de la $F_{3,2}$ es 19,16 y un F de 6 no basta; con $n = 10$ el crítico baja a 4,76 y $F = 18$ rechaza con holgura. La tercera fila hace el camino inverso: un R^2 del 1 % —un ángulo de 84°, casi ortogonalidad— resulta conjuntamente significativo con mil observaciones. Compárese con la tabla análoga de la lección 12: con $k = 2$ bastaban $n = 5$ para que $R^2 = 0,90$ fuese significativo; con $k = 4$ hacen falta más, porque el mismo R^2 conseguido con tres regresores en vez de uno es menos sorprendente —hay más direcciones sobre las que el azar puede proyectar—. Es el descuento por $k - 1$ del numerador, y es la misma idea que motivó el R^2 ajustado.

Hay una segunda advertencia, propia del F y que la lección 12 no necesitaba. El F que Gretl imprime al pie de cada tabla contrasta la significación conjunta de *todos* los regresores: la hipótesis de que ninguno de ellos tenga nada que ver con el regresando. En la práctica es la hipótesis menos interesante que cabe formular, porque nadie estima un modelo con regresores elegidos al azar; y con muestras grandes se rechaza casi sin excepción, como muestra la tercera fila de la tabla. Que ese F sea enorme no es una virtud del modelo, y que sea pequeño en una muestra grande es una señal muy seria. Los contrastes F que de verdad informan son los de la transparencia “Tres restricciones sobre **hprice2**”: nacen de una pregunta concreta —¿sobran estas dos variables?, ¿tienen el mismo efecto?, ¿vale uno la elasticidad?, ¿son constantes los rendimientos a escala?— y su respuesta cambia algo.

El F automático conviene mirarlo una vez y no volver sobre él.

La tercera advertencia es la que el laboratorio anterior ilustró con doce viviendas para el t , y que vale igual para el F . Con la misma submuestra aleatoria de doce viviendas y el modelo de **rooms**, **nox** y **stratio**, el $F(3, 8)$ de significación conjunta vale 9,55 frente a un crítico de 4,07: se rechaza H_0 , los tres regresores no son nulos a la vez. Y sin embargo el t de **stratio** es $-0,10$, el de **nox** apenas $-2,6$ y los intervalos de confianza de los tres coeficientes son enormes. El F significativo dice que *algo* del subespacio de los regresores explica el precio; no dice qué, ni cuánto, ni con qué precisión. Como en la lección 12, lo que informa es el coeficiente en las unidades del problema y su intervalo, y un F —como un t — solo dice si la evidencia contra una hipótesis es distinguible del azar. Un F no significativo tampoco es la prueba de que las restricciones sean ciertas: es la ausencia de evidencia en su contra, y con pocas observaciones esa ausencia es barata.

Y, como siempre, nada de esto establece causalidad. Un F conjunto significativo dice que la proyección de $\mathbf{y}$ sobre el subespacio de los regresores es más larga de lo que el azar produciría; de si alguno de esos regresores *causa* el regresando no dice nada, por las mismas razones de las lecciones 8, 9 y 12. Las correlaciones espurias producen F enormes con la misma facilidad que t enormes.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.2f (significación económica frente a estadística, que se aplica sin cambios al F) y sección 4.5, apartado sobre el valor p del contraste F .

10 Recapitulación y guiño a las sesiones siguientes

Pieza	Fórmula	Lectura geométrica
F de significación conjunta	$\frac{SEC/(k-1)}{SRC/(n-k)} = \frac{n-k}{k-1} \frac{R^2}{1-R^2}$	razón de catetos, cada uno por su dimensión
F de q restricciones	$\frac{(SRC_r - SRC)/q}{SRC/(n-k)}$	el mismo triángulo, con $\hat{\mathbf{y}}_r$ en el papel de $\bar{\mathbf{y}}$
Caso $q = 1$	$F = t^2, F_c = t_c^2$	el contraste t es un F de una restricción
Grados de libertad	$k - 1$ (o q) y $n - k$	dimensiones de los dos catetos
Región crítica	$F > F_c$ (una cola)	cono alrededor del subespacio de los regresores

Un único ingrediente nuevo: dividir cada cateto por su dimensión. Demostrado: Pitágoras anidado y $F = t^2$. Enunciado: la distribución $F_{q, n-k}$.

Lección 14: colinealidad. La longitud propia $\|\tilde{\mathbf{x}}_j\|_e$ de cada regresor y el *factor de inflación de la varianza* explican con números el caso de **nox** y **lnox**. *En un laboratorio posterior:* **restrict** con varias restricciones a la vez, y la distribución del F bajo H_0 con datos simulados.

Recapitulación y guiño a las sesiones siguientes

El balance de la sesión es, otra vez, mucho resultado con muy poco ingrediente nuevo. Todo el aparato —el triángulo, las sumas de cuadrados, el R^2 , la cuasivarianza $\mathfrak{s}^2$, la isotropía del ruido bajo normalidad— estaba ya sobre la mesa desde las lecciones 8, 11 y 12. Lo único que hoy hemos añadido es una idea: comparar los dos catetos *por dimensión*, dividiendo cada suma de cuadrados por el

número de direcciones del subespacio donde vive. Con ella, el contraste t de la lección 12 resulta ser el caso de una sola dirección, y el contraste de cualquier hipótesis lineal sobre los coeficientes —excluir, igualar, fijar— se reduce a una única receta: estimar el modelo con y sin la hipótesis, leer las dos sumas de cuadrados residuales y contar dimensiones.

La tabla resume las piezas. Conviene subrayar, como en la lección 12, lo que se ha demostrado y lo que no. Se ha demostrado, con material del propio curso, el Pitágoras anidado —que la subida de la suma de cuadrados residual al imponer restricciones es la longitud al cuadrado de un cateto ortogonal al residuo— y las identidades $F = \frac{n-k}{k-1} \frac{R^2}{1-R^2}$ y $F = t^2$. Se ha enunciado sin demostrar, igual que la t de Student en la lección 12, la distribución $F_{q, n-k}$ del estadístico bajo H_0 , y la afirmación que la sostiene: que un ruido isótropo reparte su longitud al cuadrado por igual entre las dimensiones. La parte geométrica es rigurosa; la parte probabilística, enunciada.

Quedan dos hilos abiertos. El primero es el caso de `nox` y `lnox`, que hoy hemos sabido *leer* pero no *medir*. La lección 12 enunció que con varios regresores $\text{ee}(\hat{\beta}_j) = \mathfrak{s}/\|\tilde{\mathbf{x}}_j\|_e$, con $\tilde{\mathbf{x}}_j$ la parte propia del regresor j ; hoy hemos visto qué pasa cuando esa parte propia es casi nula: el t individual se hunde mientras el conjunto sigue siendo significativo. La lección 14 pondrá números a $\|\tilde{\mathbf{x}}_j\|_e$, mostrará que se calcula con una regresión auxiliar de $\mathbf{x}_j$ sobre los demás regresores, y de ahí saldrá el *factor de inflación de la varianza*, la medida habitual de cuánto encarece la colinealidad cada error estándar. El segundo hilo es de laboratorio. En una práctica posterior se aplicará `restrict` con varias restricciones simultáneas sobre datos reales, comprobando estimación a estimación que el F de Gretl es el cociente de sumas de cuadrados de hoy. Y se simulará, con el `loop` del laboratorio que siguió a la lección 11, la distribución del F bajo una H_0 cierta: los histogramas mostrarán que ronda 1 y que la cola derecha es la que Gretl tabula.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (resumen del contraste F); cap. 3, sección 3.4, apartado “Multicollinearity” (con el que arranca la lección siguiente).