Índice general

Lección 14. Colinealidad: cuando los regresores casi se alinean

Marcos Bujosa

23 de septiembre de 2026

Resumen

La lección 12 dejó enunciado que el error estándar de un coeficiente es $\mathfrak{s}/\|\tilde{\boldsymbol{x}}_j\|_e$, con $\tilde{\boldsymbol{x}}_j$ la parte *propia* del regresor. Hoy definimos esa parte propia, medimos cuánto se acorta cuando un regresor casi cae en el subespacio de los demás y leemos la consecuencia: coeficientes imprecisos e inestables, con un ajuste que no se mueve.

"El subespacio es estable; las coordenadas, no."

- (slides) — (html) — (pdf)

1 De dónde venimos: el caso `nox` y `lnox`

La lección 13 dejó un caso sin explicar: en `price` sobre `rooms`, `nox` y `lnox`, los dos t de la contaminación eran nulos y el F de excluirlas juntas, 28,2.

Lo que cambia al añadir `lnox` (`hprice2`, $n = 506$):

Modelo (`price` sobre `rooms` y …)	coef. de `nox`	ee($\hat{\beta}_{\text{nox}}$)	$\mathfrak{s}$
… `nox`	−1884,7	253,6	6291,0
… `nox` y `lnox`	404,2	2198,9	6290,4

El error estándar se multiplica por 8,7 y $\mathfrak{s}$ no cambia. La lección 12 enunció $\text{ee}(\hat{\beta}_j) = \mathfrak{s}/\|\tilde{\boldsymbol{x}}_j\|_e$: lo que se ha acortado es $\tilde{\boldsymbol{x}}_{\text{nox}}$, la parte *propia* de `nox`.

Hoy: qué es $\tilde{\boldsymbol{x}}_j$, cuánto mide, qué le pasa al coeficiente y qué hacer.

De dónde venimos: el caso `nox` y `lnox`

La lección 13 terminó con un caso que supimos leer pero no medir. Con `rooms` y `nox` en el modelo, la contaminación tiene un t de −7,4; con `rooms`, `nox` y `lnox`, los t de las dos medidas de contaminación caen a 0,18 y −1,05, mientras que el F de excluirlas a la vez vale 28,2. Dijimos que era la firma de la colinealidad y que la lección 14 le pondría números. Esta es la lección 14.

La tabla de la transparencia aísla lo que ocurre. Al añadir `lnox`, el coeficiente de `nox` cambia de signo y su error estándar pasa de 253,6 a 2198,9. Lo que no cambia es $\mathfrak{s}$: 6291,0 frente a 6290,4. El ruido estimado es el mismo, así que la culpa no es del numerador del error estándar.

La lección 12 dejó enunciado, sin demostrar, que con varios regresores

$$\mathrm{ee}(\hat{\beta}_j) = \frac{\mathfrak{s}}{\|\tilde{\boldsymbol{x}}_j\|_e},$$

donde $\tilde{\boldsymbol{x}}_j$ es lo que le queda de propio al regresor j tras descontar lo que comparte con los demás. Si $\mathfrak{s}$ no ha cambiado y el error estándar se ha multiplicado por 8,7, la parte propia de `nox` se ha dividido por 8,7. Eso es lo que hay que entender: qué es exactamente $\tilde{\boldsymbol{x}}_j$, por qué se acorta y cuánto.

La sesión tiene tres partes. Las cuatro primeras transparencias definen la parte propia de un regresor y miden su longitud, con el único resultado que hoy se demuestra. Las dos siguientes ponen números al caso `nox` y `lnox` y explican por qué los coeficientes oscilan. Las tres últimas son de lectura: síntomas, qué es y qué no es la colinealidad, y qué hacer.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4, apartado "Multicollinearity" (el planteamiento del problema con la fórmula de la varianza de $\hat{\beta}_j$).

2 Colinealidad exacta y colinealidad de grado

Exacta (lección 10): un regresor es combinación lineal de los demás. $\mathbf{X}^\top\mathbf{X}$ no es invertible; infinitos $\widehat{\boldsymbol{\beta}}$ dan el mismo ajuste. Ejemplo: temperatura en Celsius y en Fahrenheit.

De grado (hoy): un regresor es *casi* combinación lineal de los demás. Solución única, pero imprecisa e inestable.

Con dos regresores no constantes: $\rho_{\boldsymbol{x}_2\boldsymbol{x}_3} \to \pm 1$, vectores en desviaciones casi alineados (lección 3: $\cos\theta \to \pm 1$). Con más: alguno casi cae en el subespacio de los otros, y las correlaciones por parejas no bastan para verlo (lección 10).

No es una frontera: es una escala. La pregunta útil no es "¿hay colinealidad?" sino "¿cuánto le cuesta a cada coeficiente?".

Colinealidad exacta y colinealidad de grado

El curso ya conoce el caso extremo. En la lección 10, al deducir las ecuaciones normales, dijimos que si los regresores son linealmente dependientes la matriz $\mathbf{X}^\top\mathbf{X}$ no es invertible y el sistema tiene infinitas soluciones: infinitos vectores $\widehat{\boldsymbol{\beta}}$ producen el mismo ajuste $\widehat{\boldsymbol{y}}$. Lo llamamos *colinealidad exacta*, y la práctica que siguió a esa lección lo comprobó con la temperatura en grados Celsius y en grados Fahrenheit: Gretl omite una de las dos columnas y avisa. Es una situación que se detecta sola, porque no hay nada que estimar.

El caso de hoy es el vecino: un regresor no es combinación lineal exacta de los demás, pero casi. La solución existe y es única, pero el coeficiente de ese regresor se estima con poca precisión y cambia mucho ante cambios pequeños de los datos. Lo llamaremos *colinealidad de grado*, y la palabra "grado" es la importante: no es una propiedad que se tenga o no se tenga, sino una escala continua, y lo que hay que medir es cuánto cuesta.

Con dos regresores no constantes la escala es la correlación. La lección 5 definió $\rho_{\boldsymbol{x}_2\boldsymbol{x}_3}$ como el coseno del ángulo entre los vectores en desviaciones, y la lección 3 mostró que $\cos\theta = \pm 1$ exactamente cuando los vectores están alineados. Colinealidad exacta es $\rho = \pm 1$; colinealidad de

grado es ρ cerca de ± 1, vectores casi alineados. La lección 3 lo anunció con estas palabras: dos regresores muy correlacionados son vectores casi alineados.

Con tres o más regresores no constantes la correlación por parejas deja de bastar, por la razón que dio la lección 10: un regresor puede estar cerca del subespacio generado por los demás sin estar cerca de ninguno de ellos en particular. El ejemplo de allí sirve: la renta total de un hogar frente a sus dos componentes, renta del trabajo y renta del capital. Las tres transparencias siguientes construyen la medida que sí sirve en todos los casos: la longitud de la parte propia de cada regresor.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.3 (supuesto MLR.3, no colinealidad perfecta) y sección 3.4 (la distinción con la colinealidad "imperfecta").

3 La parte propia de un regresor

Proyectamos $\boldsymbol{x}_j$ sobre el subespacio de *los demás* regresores (constante incluida), como en la lección 10. El residuo de esa *regresión auxiliar* es la parte propia:

$$\boldsymbol{x}_j = \underbrace{(\text{combinación de los demás})}_{\text{lo que comparte}} + \underbrace{\tilde{\boldsymbol{x}}_j}_{\text{lo propio}}, \qquad \tilde{\boldsymbol{x}}_j \perp \text{ todos los demás regresores.}$$

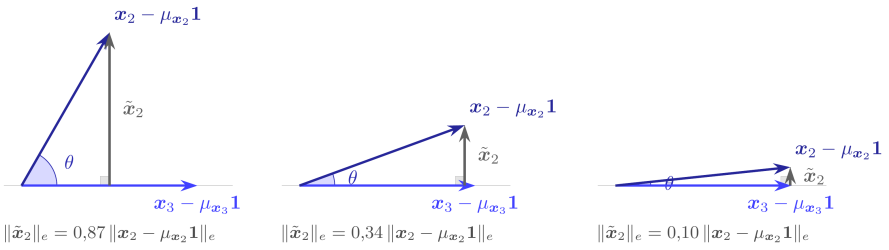

Figura 1: Dos regresores en desviaciones en el plano $\mathcal{L}(\mathbf{1})^{\perp}$, con ángulo θ de 60°, 20° y 6°. En azul, $\boldsymbol{x}_2 - \mu_{\boldsymbol{x}_2}\mathbf{1}$ y $\boldsymbol{x}_3 - \mu_{\boldsymbol{x}_3}\mathbf{1}$; la línea de puntos baja desde la punta del primero hasta la recta del segundo, y el vector gris que sube desde ese pie es $\tilde{\boldsymbol{x}}_2$, la parte propia de $\boldsymbol{x}_2$. La longitud del regresor es la misma en los tres paneles; la de su parte propia es $\operatorname{sin}\theta$ veces esa longitud: 0,87, 0,34 y 0,10.

Enunciado en la lección 12 (sigue sin demostrarse): $\text{ee}(\hat{\beta}_j) = \mathfrak{s}/\|\tilde{\boldsymbol{x}}_j\|_e$. Con un solo regresor no constante, $\tilde{\boldsymbol{x}}_2 = \boldsymbol{x}_2 - \mu_{\boldsymbol{x}_2}\mathbf{1}$ y se recupera la lección 12.

La parte propia de un regresor

La lección 12 habló de "la parte del regresor j que no se puede reconstruir a partir de los demás". Hoy le damos definición, y la definición no tiene nada nuevo: es una proyección ortogonal de las de la lección 10, aplicada a un regresor en lugar de al regresando.

Tomemos el regresor $\boldsymbol{x}_j$ y el subespacio generado por todos los demás, constante incluida: $\mathcal{L}(\mathbf{1}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_k)$. Proyectar $\boldsymbol{x}_j$ sobre ese subespacio es hacer una regresión, la *regresión auxiliar* de $\boldsymbol{x}_j$ sobre los demás regresores, y descompone $\boldsymbol{x}_j$ en dos piezas ortogonales: el

ajuste de esa regresión, que es la parte de $\boldsymbol{x}_j$ que los demás regresores reproducen, y el residuo, que es la parte que no reproducen. A ese residuo lo llamamos *parte propia* del regresor y lo denotamos $\tilde{\boldsymbol{x}}_j$. Por ser un residuo de la lección 10, $\tilde{\boldsymbol{x}}_j$ es ortogonal a todos los generadores del subespacio: a $\mathbf{1}$ y a cada uno de los otros regresores. Tiene, en particular, media cero.

La figura lo muestra con dos regresores no constantes, en el plano de las desviaciones. Si solo hay otro regresor, $\boldsymbol{x}_3$, el subespacio de "los demás" en desviaciones es la recta de $\boldsymbol{x}_3 - \mu_{\boldsymbol{x}_3}\mathbf{1}$, y la parte propia de $\boldsymbol{x}_2$ es la componente de $\boldsymbol{x}_2 - \mu_{\boldsymbol{x}_2}\mathbf{1}$ perpendicular a esa recta: el cateto vertical del triángulo. Los tres paneles tienen el mismo regresor, con la misma longitud; solo cambia el ángulo θ con el otro. Con 60° la parte propia es casi todo el regresor; con 6° es una décima parte. La transparencia siguiente pone la fórmula.

Con esta definición, el enunciado de la lección 12 se lee así: el error estándar de $\hat{\beta}_j$ es el ruido estimado por unidad de longitud *de la parte propia* del regresor. Es la misma lectura de la lección 12, "ruido por unidad de longitud del regresor en desviaciones", con una corrección: con varios regresores no cuenta toda la longitud del regresor, solo la que no comparte con los demás. Cuando solo hay un regresor no constante no hay nada que compartir, la regresión auxiliar es la de $\boldsymbol{x}_2$ sobre $\mathbf{1}$, su residuo es el vector en desviaciones $\boldsymbol{x}_2 - \mu_{\boldsymbol{x}_2}\mathbf{1}$, y la fórmula es la de la lección 12. El enunciado sigue sin demostrarse: su prueba necesita el resultado de que $\hat{\beta}_j$ puede obtenerse proyectando $\boldsymbol{y}$ sobre la dirección de $\tilde{\boldsymbol{x}}_j$, que este curso no desarrolla. Lo que sí vamos a demostrar es cuánto mide $\tilde{\boldsymbol{x}}_j$.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.2, apartado "A partialling out interpretation of multiple regression" (la regresión auxiliar y su residuo, con el mismo papel que aquí).

4 Cuánto se acorta: el factor de inflación de la varianza

Pitágoras en la regresión auxiliar (lección 8), con R_j^2 su coeficiente de determinación:

$$\|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e^2 = \mathrm{SEC}_j + \|\tilde{\boldsymbol{x}}_j\|_e^2 \implies \boxed{\|\tilde{\boldsymbol{x}}_j\|_e = \|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e \sqrt{1 - R_j^2} = \|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e \operatorname{sin}\theta_j.}$$

Sustituyendo en el error estándar:

$$\mathrm{ee}(\hat{\beta}_j) = \frac{\mathfrak{s}}{\|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e} \cdot \frac{1}{\sqrt{1 - R_j^2}} = \underbrace{\frac{\mathfrak{s}}{\|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e}}_{\text{como si estuviera solo}} \cdot \sqrt{\mathrm{VIF}_j}, \qquad \mathrm{VIF}_j = \frac{1}{1 - R_j^2} = \frac{1}{\operatorname{sin}^2\theta_j}.$$

Con $k = 3$, $R_2^2 = \rho_{\boldsymbol{x}_2\boldsymbol{x}_3}^2$:

$\rho_{\boldsymbol{x}_2\boldsymbol{x}_3}$	θ	VIF	$\sqrt{\mathrm{VIF}}$
0,5	60°	1,3	1,2
0,9	26°	5,3	2,3
0,99	8°	50	7,1
0,999	2,6°	500	22

Cuánto se acorta: el factor de inflación de la varianza

Este es el único resultado que la sesión demuestra, y sale de la lección 8 sin ningún ingrediente nuevo. La regresión auxiliar de $\boldsymbol{x}_j$ sobre los demás regresores es una regresión como cualquier otra, con $\boldsymbol{x}_j$ en el papel del regresando, así que tiene su triángulo rectángulo y su identidad STC = SEC + SRC. Escribamos las tres sumas con su subíndice j para recordar de qué regresión son. La suma total es la longitud al cuadrado del regresor en desviaciones, $\mathrm{STC}_j = \|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e^2$. La suma residual es la longitud al cuadrado de la parte propia, $\mathrm{SRC}_j = \|\tilde{\boldsymbol{x}}_j\|_e^2$, porque $\tilde{\boldsymbol{x}}_j$ es el residuo de esa regresión. Y el coeficiente de determinación de la regresión auxiliar, que llamaremos R_j^2, es $\mathrm{SEC}_j/\mathrm{STC}_j = 1 - \mathrm{SRC}_j/\mathrm{STC}_j$. Despejando,

$$\|\tilde{\boldsymbol{x}}_j\|_e^2 = \mathrm{SRC}_j = \mathrm{STC}_j\,(1 - R_j^2) = \|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e^2\,(1 - R_j^2).$$

Tomando raíces queda la primera igualdad del recuadro. La segunda es la lección 8 otra vez: $R_j^2 = \cos^2\theta_j$, con θ_j el ángulo entre el regresor en desviaciones y su ajuste en desviaciones en la regresión auxiliar, así que $1 - R_j^2 = \sin^2\theta_j$ y $\|\tilde{\boldsymbol{x}}_j\|_e = \|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e \sin\theta_j$. Es lo que la figura de la transparencia anterior muestra: el cateto vertical mide la hipotenusa por el seno del ángulo. ■

Ahora sustituimos en el enunciado de la lección 12. El error estándar queda escrito como producto de dos factores. El primero, $\mathfrak{s}/\|\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\|_e$, es el error estándar que tendría $\hat{\beta}_j$ si el regresor no compartiera nada con los demás: la fórmula de la lección 12, con el $\mathfrak{s}$ del modelo completo. El segundo, $1/\sqrt{1 - R_j^2}$, es el precio de compartir. Su cuadrado tiene nombre propio en los manuales: *factor de inflación de la varianza*, $\mathrm{VIF}_j = 1/(1 - R_j^2)$, porque multiplica la varianza del estimador; al error estándar lo multiplica su raíz. En el lenguaje del curso, $\mathrm{VIF}_j = 1/\sin^2\theta_j$: cuanto más se cierra el ángulo entre un regresor y el subespacio de los demás, más se infla.

Con dos regresores no constantes la regresión auxiliar es simple, $\boldsymbol{x}_2$ sobre $\mathbf{1}$ y $\boldsymbol{x}_3$, y la lección 8 demostró que en regresión simple $R^2 = \rho^2$. Por tanto $R_2^2 = \rho^2_{\boldsymbol{x}_2\boldsymbol{x}_3}$ y $\mathrm{VIF}_2 = 1/(1 - \rho^2_{\boldsymbol{x}_2\boldsymbol{x}_3})$, y lo mismo para $\boldsymbol{x}_3$: con $k = 3$ los dos regresores tienen el mismo VIF. La tabla de la transparencia recorre la escala. Una correlación de 0,5 cuesta poco: el error estándar crece un 15 %. Una de 0,9 lo multiplica por 2,3. Una de 0,99, por 7; y con 0,999 el ángulo es de 2,6° y el error estándar se multiplica por 22. El crecimiento no es lineal en ρ: casi todo el coste se paga al final, en los últimos grados antes de la alineación exacta. Es la razón de que la colinealidad pase desapercibida hasta que de pronto no lo hace.

Con más de dos regresores no constantes no hay atajo: R_j^2 es el R^2 de una regresión múltiple, y hay que estimarla. Gretl lo hace por nosotros con un comando, como veremos en la transparencia siguiente.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4, ecuación de $\mathrm{Var}(\hat{\beta}_j)$ con el factor $1 - R_j^2$ y la discusión del VIF que la sigue.

5 El caso `nox` y `lnox` con números

Regresión auxiliar de `nox` sobre `rooms` y `lnox` (`hprice2`): $R_j^2 = 0{,}9879$.

Cantidad	Valor
$\|\boldsymbol{x}_{\text{nox}} - \mu_{\text{nox}}\mathbf{1}\|_e$ (longitud en desviaciones)	26,03
$\sqrt{1 - R_j^2} = \sin\theta_j$	0,1099
$\|\tilde{\boldsymbol{x}}_{\text{nox}}\|_e$ (parte propia)	2,861
$\text{VIF} = 1/(1 - R_j^2)$; $\sqrt{\text{VIF}}$	82,8; 9,10
$\mathfrak{s}/\|\tilde{\boldsymbol{x}}_{\text{nox}}\|_e = 6290{,}37/2{,}861$	2198,9

Es exactamente el error estándar de la tabla de Gretl. Sin `lnox`, la parte propia de `nox` (sobre **1** y `rooms`) medía 24,81: entrar `lnox` la divide por 8,7.

En Gretl: comando `vif` tras estimar, o `Análisis -> Colinealidad` en la ventana del modelo. Devuelve 82,8 para `nox`, 82,9 para `lnox` y 1,1 para `rooms`.

El caso `nox` y `lnox` con números

Volvamos al modelo con que abrimos la sesión y hagamos la cuenta completa para `nox`. La regresión auxiliar es la de `nox` sobre la constante, `rooms` y `lnox`, y su coeficiente de determinación es $R_j^2 = 0{,}9879$: los otros regresores reproducen el 98,8 % de la variación de `nox`. Casi todo, como cabía esperar de una variable y su logaritmo, cuya correlación es 0,9939, un ángulo de 6,3°.

Las longitudes salen de la fórmula de la transparencia anterior. El regresor en desviaciones mide 26,03; $\sqrt{1 - R_j^2} = 0{,}1099$; la parte propia mide $26{,}03 \cdot 0{,}1099 = 2{,}861$. Y el error estándar es $\mathfrak{s}$ dividido por esa longitud: $6290{,}37/2{,}861 = 2198{,}9$, la cifra que Gretl imprime en la tabla de la lección 13. El enunciado de la lección 12, que no hemos demostrado, se cumple aquí con cuatro cifras.

Conviene leer también el factor por el que se ha multiplicado el error estándar. En el modelo sin `lnox`, la parte propia de `nox` es su residuo sobre **1** y `rooms`, y mide 24,81: casi toda su longitud en desviaciones, porque `rooms` apenas comparte nada con `nox`. Al entrar `lnox`, la parte propia baja a 2,861, un factor 8,7, y el error estándar sube en el mismo factor, de 253,6 a 2198,9. No ha cambiado el ruido ni el número de observaciones; solo la longitud con la que `nox` se mide a sí misma.

Gretl calcula los factores de inflación con el comando `vif`, que se ejecuta después de estimar el modelo, o desde el menú `Análisis -> Colinealidad` de la ventana del modelo. Para nuestro modelo devuelve 82,8 para `nox`, 82,9 para `lnox` y 1,1 para `rooms`: las dos medidas de contaminación se estorban entre sí y ninguna estorba a `rooms`. Con la salida de Gretl y esta lección, un VIF se lee sin misterio: es $1/\sin^2\theta_j$, y su raíz es el factor por el que se ha multiplicado el error estándar respecto al que el regresor tendría si no compartiera nada. Con 82,8, ese factor es 9,10; y en efecto $241{,}6 \cdot 9{,}10 = 2198{,}9$, donde 241,6 es $\mathfrak{s}$ dividido por la longitud completa del regresor en desviaciones.

Para profundizar: manual de Gretl, comando `vif`; Wooldridge, J. M. (2020), cap. 3, sección 3.4 (el VIF como estadístico de diagnóstico y sus límites).

6 Por qué oscilan los coeficientes

$\widehat{\boldsymbol{y}}$ es la proyección sobre el subespacio: estable. $\hat{\beta}_2, \hat{\beta}_3$ son sus *coordenadas* respecto de los generadores: con generadores casi alineados, inestables.

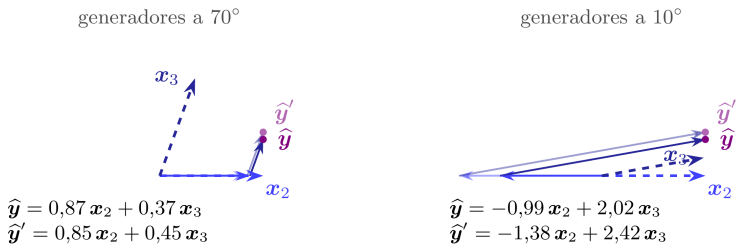

Figura 2: El mismo ajuste $\widehat{\boldsymbol{y}}$ escrito como $\hat{\beta}_2\boldsymbol{x}_2 + \hat{\beta}_3\boldsymbol{x}_3$, con los dos regresores en desviaciones y de longitud 1 (azul, a trazos). Un segundo ajuste $\widehat{\boldsymbol{y}}'$ dista 0,07 del primero. A la izquierda los generadores forman 70° y las coordenadas apenas cambian; a la derecha forman 10° y cambian 0,40 cada una, en sentidos opuestos: casi seis veces el desplazamiento del ajuste, que es $1/\operatorname{sin} 10°$. Los tramos azules muestran las dos descomposiciones (más claros los de $\widehat{\boldsymbol{y}}'$).

Un desplazamiento de $\widehat{\boldsymbol{y}}$ perpendicular a $\boldsymbol{x}_2$ mueve las coordenadas $1/\operatorname{sin}\theta = \sqrt{\text{VIF}}$ veces más, y en sentidos opuestos: $\hat{\beta}_2$ y $\hat{\beta}_3$ se compensan. SRC, R^2 y F no se enteran.

Por qué oscilan los coeficientes

El error estándar dice cuánto se dispersa un coeficiente de una muestra a otra. La figura dice por qué se dispersa tanto cuando los regresores casi se alinean, y lo dice sin probabilidad: es geometría de coordenadas.

Recordemos la lección 10: $\widehat{\boldsymbol{y}}$ es la proyección de $\boldsymbol{y}$ sobre el subespacio de los regresores, y $\widehat{\boldsymbol{\beta}}$ son las coordenadas de esa proyección respecto de los generadores, lo que hay que avanzar en la dirección de cada regresor para llegar a $\widehat{\boldsymbol{y}}$. La proyección depende del subespacio, no de qué generadores lo describan. Las coordenadas sí dependen de los generadores, y de cómo estén colocados.

La figura toma dos regresores en desviaciones de longitud 1 y un ajuste $\widehat{\boldsymbol{y}}$ en el plano que generan. Cambiar la muestra cambia un poco el ajuste; lo representamos con un segundo ajuste $\widehat{\boldsymbol{y}}'$ a distancia 0,07 del primero. Con los generadores a 70°, las coordenadas pasan de (0,87; 0,37) a (0,85; 0,45): cambios del mismo orden que el desplazamiento. Con los generadores a 10°, pasan de (−0,99; 2,02) a (−1,38; 2,42): cada coordenada cambia 0,40, casi seis veces el desplazamiento, y una sube lo que la otra baja. La cuenta exacta es sencilla para un desplazamiento perpendicular a $\boldsymbol{x}_2$: la coordenada de $\boldsymbol{x}_3$ cambia $1/\operatorname{sin}\theta$ veces el desplazamiento, y la de $\boldsymbol{x}_2$ lo compensa con signo contrario. Ese factor $1/\operatorname{sin}\theta$ es $\sqrt{\text{VIF}}$, el mismo de la transparencia anterior, y no es casualidad: el error estándar mide con probabilidad lo que la figura muestra con geometría.

Dos consecuencias. La primera es que los coeficientes de regresores casi alineados no solo son imprecisos, sino que lo son *juntos y al revés*: cuando el azar de la muestra empuja uno hacia arriba, empuja al otro hacia abajo. La lección 11 calculó, para $k = 2$, la covarianza entre $\hat{\beta}_1$ y $\hat{\beta}_2$, negativa cuando la media del regresor es positiva; con regresores alineados, la covarianza entre $\hat{\beta}_2$ y $\hat{\beta}_3$ es muy negativa, y así lo comprobaremos en el laboratorio de colinealidad con un experimento de Monte Carlo. Es lo que se veía en la tabla de la lección 13: `nox` con coeficiente 404 y `lnox` con −13262,

dos cifras enormes de signo opuesto que se reparten un efecto que, sumado, es el de siempre.

La segunda consecuencia es la que hay que retener para leer resultados. Lo que la colinealidad estropea son las coordenadas; no el ajuste. $\widehat{\boldsymbol{y}}$ es el mismo punto se describa como se describa, así que SRC, R^2, $\mathfrak{s}$ y el F de significación conjunta no cambian: en la tabla de apertura, $\mathfrak{s}$ vale lo mismo con y sin `lnox`, y el R^2 pasa de 0,5352 a 0,5362. Por eso la lección 13 podía decir que las dos variables juntas explican el precio aunque ninguna lo haga por separado: el subespacio que generan es estable; el reparto del efecto entre ellas, no.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4 (el ejemplo de gasto por alumno en varias partidas, en el que dos regresores casi alineados tienen coeficientes imprecisos mientras el conjunto está bien estimado).

7 Síntomas y diagnóstico

- F conjunto alto con t individuales bajos (lección 13), o R^2 alto con casi nada significativo.
- Errores estándar que se disparan al añadir un regresor, sin que cambie $\mathfrak{s}$.
- Coeficientes muy sensibles a quitar unas pocas observaciones o a añadir un regresor.
- Correlaciones altas entre regresores; pero con $k > 3$ no bastan: el diagnóstico es el VIF_j de cada uno.

Una distinción: que un coeficiente cambie al añadir un regresor es normal, es el efecto parcial de la lección 10. Colinealidad es que su error estándar se dispare y que cambie mucho con poca información nueva.

Qué mirar: $\sqrt{\text{VIF}_j}$, el factor por el que se ha multiplicado el error estándar de $\hat{\beta}_j$. Es una escala, no un veredicto.

Síntomas y diagnóstico

La colinealidad de grado no salta a la vista como la exacta, porque Gretl estima el modelo sin protestar. Hay que reconocerla en la tabla, y los síntomas son los que las lecciones anteriores ya han producido.

El primero es el de la lección 13: un F de significación conjunta alto con t individuales bajos, o un R^2 alto en un modelo donde casi ningún coeficiente es significativo. El F y el R^2 miran el subespacio; los t miran las coordenadas. Cuando los dos veredictos discrepan, es que el subespacio explica y las coordenadas no se dejan separar.

El segundo es el de la tabla de apertura: un error estándar que se dispara al añadir un regresor mientras $\mathfrak{s}$ no cambia. Conviene contrastarlo con lo que sí es normal. En la lección 10 y en su laboratorio vimos que un coeficiente cambia de valor, y hasta de signo, al añadir un regresor: es el efecto parcial, el significado del coeficiente ha cambiado porque se mantiene fija otra variable. Eso no es colinealidad. Colinealidad es que, además, el error estándar se multiplique por un factor grande, y que el coeficiente se vuelva sensible a cambios pequeños de la muestra: quitar unas pocas observaciones y verlo saltar, como ocurría con las viviendas del laboratorio de regresión múltiple.

El tercero es la correlación entre regresores, el síntoma más popular y el menos fiable. Con dos regresores no constantes basta, porque VIF $= 1/(1-\rho^2)$. Con más, la lección 10 explicó por qué no: un regresor puede ser casi combinación de otros dos sin estar muy correlado con ninguno. La medida que sirve en todos los casos es el VIF de cada regresor, que Gretl calcula con `vif`, y en particular su raíz, que es el factor por el que se ha multiplicado el error estándar.

Los manuales suelen dar una regla: un VIF mayor que 10 indica un problema. Conviene saberla, porque Gretl la imprime al pie de la salida de `vif`, y conviene saber también que es arbitraria: 10 significa que el error estándar se ha multiplicado por 3,2, y si ese factor es grave o no depende de para qué se quiera el coeficiente. Un VIF de 5 que convierte un intervalo estrecho en uno inservible es un problema; un VIF de 50 en un coeficiente que sigue siendo significativo y con un intervalo aceptable no lo es. Léase la escala, no el umbral.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4 (por qué la regla del VIF mayor que 10 es arbitraria y qué debe mirarse en su lugar).

8 Qué es y qué no es

No es una violación de los supuestos: MCO sigue insesgado y BLUE (lección 11) y los contrastes siguen siendo válidos (lección 12). Los errores estándar grandes son correctos: dicen la verdad sobre la muestra.

No estropea el ajuste, el R^2 ni la predicción: $\widehat{\boldsymbol{y}}$ es el mismo.

Es una propiedad de la muestra: falta variación *propia*. Es la misma enfermedad que tener pocos datos.

Para estimar el efecto de una variable hace falta que esa variable varíe *por su cuenta*.

Qué es y qué no es

Los manuales tratan la colinealidad en el capítulo de "problemas" del modelo, y conviene precisar qué clase de problema es, porque no es del tipo de los que vendrán en las lecciones siguientes.

No es una violación de los supuestos. Los supuestos del Modelo Clásico hablan de la perturbación y del modelo poblacional; la colinealidad de grado habla de los regresores de la muestra, y ningún supuesto prohíbe que estén correlados. El único que los menciona, la independencia lineal de la lección 10, prohíbe solo el caso exacto. Con colinealidad de grado, todo lo demostrado sigue en pie: $\hat{\beta}_j$ es insesgado, su varianza es la que dice la fórmula y no hay estimador lineal insesgado con menos (lección 11); los errores estándar de Gretl son correctos y los contrastes t y F valen (lección 12). Los errores estándar grandes no son un fallo del método: son la información verdadera de que esa muestra no permite estimar ese coeficiente con precisión.

Tampoco estropea el ajuste. La transparencia anterior lo dejó claro: $\widehat{\boldsymbol{y}}$ es la proyección sobre el subespacio, y el subespacio no sabe cómo se le describe. SRC, R^2, $\mathfrak{s}$ y las predicciones son los mismos con generadores alineados que con generadores perpendiculares. Si lo que se quiere del modelo es predecir, la colinealidad es indiferente.

Lo que sí es: una propiedad de la muestra, la falta de variación propia. Un regresor con parte

propia corta es un regresor que, dentro de esta muestra, casi no se mueve por su cuenta; y sin movimiento propio no hay forma de medir su efecto separado. Es exactamente la misma situación que tener pocos datos, y conviene decirlo así: con doce viviendas el intervalo de `rooms` iba de 1900 a 20000 dólares porque el regresor en desviaciones era corto; con `nox` y `lnox` juntas, la parte propia de cada una es corta. En los dos casos falta longitud, y por eso los dos remedios de la transparencia siguiente son el mismo: más información.

Este es el punto de llegada de una idea que el curso repite desde la lección 11. Allí, la fórmula de la varianza decía "más dispersión en X, más precisión". La lección 12 la dibujó con el regresor concentrado y el disperso: para estimar bien el efecto de una variable hace falta que esa variable varíe. Hoy, con varios regresores, la idea toma su forma definitiva: hace falta que varíe *por su cuenta*, no a la vez que las demás.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4, la comparación entre colinealidad y tamaño muestral pequeño; Goldberger, A. S. (1991), *A Course in Econometrics*, cap. 23, donde acuñó "micronumerosidad" para subrayar que son el mismo problema.

9 Qué hacer y qué no

Más información: más observaciones, o datos en los que los regresores varíen por separado (en un experimento, dosis separadas).

Reformular: si dos variables miden lo mismo, quedarse con una o usar una combinación (`nox` o `lnox`: cualquiera de los dos modelos de la lección 13). Si la teoría da una restricción, imponerla y contrastarla con el F (lección 13).

Reconocerlo: si la pregunta exige separar los efectos y la muestra no lo permite, decirlo e informar los intervalos.

Lo que no: eliminar los dos regresores "porque no son significativos" (lección 13), ni la eliminación secuencial por valor p. Fuera del curso: regresión *ridge* y componentes principales.

Qué hacer y qué no

Si la colinealidad es falta de variación propia, el remedio de fondo es más variación propia, y hay tres maneras de conseguirla o de renunciar a ella con honradez.

La primera es más información. Más observaciones alargan los regresores en desviaciones, y con ellos sus partes propias, aunque la correlación no cambie: el VIF es el mismo, pero el error estándar "como si estuviera solo" baja. Mejor aún es una muestra en la que los regresores varíen por separado. Quien diseña un experimento puede elegir las dosis, y la lección práctica del laboratorio de la lección 11 se aplica dos veces: dosis separadas para cada factor, y combinaciones de dosis que no vayan siempre juntas. Con datos económicos observados no se elige, y esa es la razón de que la colinealidad sea tan habitual en economía aplicada.

La segunda es reformular el modelo, y aquí hay que distinguir dos casos. Si dos variables miden la misma cosa, tener las dos no aporta información y quita precisión: `nox` y `lnox` son la misma contaminación en dos escalas, y la lección 13 mostró que cualquiera de los dos modelos con una sola es excelente. Quedarse con una, o usar una combinación de ambas, no es una pérdida. El otro

caso es que la teoría diga algo sobre los coeficientes: si dos efectos deben ser iguales, o sumar uno, imponer esa restricción funde regresores y elimina la dirección compartida. La lección 13 enseñó a imponerla y a contrastarla con el F; si el F no la rechaza, el modelo restringido tiene menos colinealidad y una interpretación más limpia.

La tercera es reconocerlo. Si la pregunta económica exige separar el efecto de `nox` del de `lnox`, estos datos no lo permiten, y lo correcto es decirlo: informar los dos coeficientes con sus intervalos, que serán amplios, y no fingir una precisión que la muestra no da. Un intervalo ancho es información sobre lo que los datos no saben.

Lo que no hay que hacer lo dijo la lección 13: eliminar los dos regresores porque sus t son bajos, cuando el F dice que juntos explican. Tampoco hay que fiarse de los procedimientos que eliminan regresores uno a uno según su valor p: con regresores alineados, el primero que sale lo hace por azar, y el que queda hereda el efecto de los dos. Existen, por último, métodos que renuncian a la insesgadez a cambio de varianza, como la regresión *ridge*, o que sustituyen los regresores por combinaciones ortogonales, como las componentes principales. Quedan fuera de este curso; baste saber que existen y que ninguno crea la información que la muestra no tiene.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4 (qué hacer ante la colinealidad, y por qué eliminar regresores puede ser peor que dejarlos); Kennedy, P. (2008), *A Guide to Econometrics*, cap. 12 (un repaso claro de los remedios y de sus costes).

10 Recapitulación y guiño a las sesiones siguientes

Pieza	Fórmula	Lectura geométrica
Parte propia	$\tilde{\boldsymbol{x}}_j$: residuo de $\boldsymbol{x}_j$ sobre los demás	lo que $\boldsymbol{x}_j$ no comparte
Su longitud	$\lVert\tilde{\boldsymbol{x}}_j\rVert_e = \lVert\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\rVert_e \operatorname{sen}\theta_j$	cateto perpendicular al subespacio de los demás
Error estándar	$\mathfrak{s}/\lVert\tilde{\boldsymbol{x}}_j\rVert_e$ (enunciado)	ruido por unidad de longitud propia
VIF	$1/(1-R_j^2) = 1/\operatorname{sen}^2\theta_j$	inflación de la varianza; su raíz, del error estándar
Coeficientes	coordenadas de $\widehat{\boldsymbol{y}}$	inestables si los generadores casi se alinean

Demostrado: la longitud de la parte propia. Enunciado: el error estándar. Ningún supuesto nuevo, ningún supuesto violado.

Laboratorios: el F con datos reales y simulados; y la colinealidad creciente en Monte Carlo, la regresión auxiliar de `nox` y el VIF. *Lección 16*: la trampa de las variables ficticias es colinealidad exacta con $\mathbf{1}$.

Recapitulación y guiño a las sesiones siguientes

La sesión ha añadido una definición y un resultado. La definición es la parte propia de un regresor, el residuo de proyectarlo sobre los demás: una regresión de la lección 10 con un regresor en el papel del regresando. El resultado es su longitud, $\lVert\boldsymbol{x}_j - \mu_{\boldsymbol{x}_j}\mathbf{1}\rVert_e \operatorname{sen}\theta_j$, que sale de Pitágoras en esa regresión auxiliar. Todo lo demás es lectura: el error estándar enunciado en la lección 12

dividido por esa longitud, el VIF como inverso del seno al cuadrado, y las coordenadas que oscilan porque los generadores casi se alinean.

Conviene retener también lo que la sesión no ha hecho. No ha añadido ningún supuesto, y ninguno de los anteriores se ha violado: la colinealidad de grado es compatible con todo lo demostrado en las lecciones 11 y 12. Es un problema de la muestra, la falta de variación propia, y su remedio es información, no otro estimador.

Quedan tres hilos. Dos son de laboratorio. En la práctica del contraste F se aplicará lo de la lección 13 con datos reales y se simulará la distribución del F bajo la hipótesis nula. En la práctica de colinealidad se verá, con un experimento de Monte Carlo, cómo crece la dispersión de los coeficientes al cerrar el ángulo entre dos regresores mientras el ajuste no se mueve; se calculará a mano la parte propia de `nox` con la regresión auxiliar; y se comprobará el VIF de Gretl contra la fórmula de hoy. El tercer hilo es de teoría y llega en la lección 16, al introducir las variables ficticias: si se incluyen una indicadora de hombre y otra de mujer junto con la constante, la suma de las dos es $\mathbf{1}$, y eso es colinealidad exacta de la lección 10. La llamada "trampa de las variables ficticias" no es una regla nueva: es la que hoy hemos repasado, en su caso extremo.

Para profundizar: Wooldridge, J. M. (2020), cap. 3, sección 3.4 (resumen) y cap. 7, sección 7.2 (la trampa de las variables ficticias, con la que enlaza la lección 16).