Índice general

Lección 12. Inferencia: error estándar, intervalo de confianza y contraste t

Marcos Bujosa

19 de septiembre de 2026

Resumen

La lección 11 dejó a $\hat{\beta}_2$ bien centrado y con una varianza conocida, pero le faltaba lo esencial para contrastar hipótesis: la *forma* de su distribución. Hoy la obtenemos suponiendo normalidad de la perturbación.

Con ella construimos el error estándar, el intervalo de confianza y el contraste t —que no mide nada que el R^2 no midiera ya— y aprendemos a *leer* una tabla de resultados.

"Significativo" no quiere decir "grande", ni "importante", ni "verdadero".

- (slides) — (html) — (pdf)

1 De dónde venimos: nos falta la forma

De la lección 11 nos traemos dos resultados sobre el modelo lineal simple ($k = 2$: la constante *1* y un único regresor no constante X):

$$\mathrm{E}(\hat{\beta}_2) = \beta_2, \qquad Var[\hat{\beta}_2 \mid \boldsymbol{X}] = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2} = \frac{\sigma^2}{nS^2}.$$

Y nos traemos también una limitación reconocida en voz alta: eso *no basta* para responder a

- ¿es compatible con estos datos que β_2 valga cero?
- ¿en qué rango de valores puedo confiar razonablemente que está β_2?

Saber dónde está centrada una distribución y cuánto se dispersa no dice cuánta probabilidad hay *más allá* de un punto. Para eso hace falta la **forma**.

Hoy: un supuesto nuevo, tres herramientas (error estándar, intervalo, contraste) — y, sobre todo, aprender a *leerlas*.

De dónde venimos: nos falta la forma

La lección 11 terminó con un reconocimiento explícito de sus propios límites, y merece la pena recordar por qué era un límite real y no una coquetería. Allí demostramos que $\hat{\beta}_2$ está bien centrado en β_2 (insesgadez) y calculamos exactamente cuánto se dispersa alrededor de ese centro

($Var[\hat{\beta}_2 \mid \boldsymbol{X}]$). Con esas dos cantidades sabemos mucho, pero no sabemos lo único que permite hacer *inferencia*: cuánta probabilidad acumula la distribución de $\hat{\beta}_2$ a partir de un valor dado.

Un ejemplo elemental lo deja claro. Imagine dos variables aleatorias con la misma media, 0, y la misma varianza, 1. La primera es normal estándar; la segunda toma solo los valores -1 y $+1$, con probabilidad $1/2$ cada uno. Ambas tienen idénticos los dos primeros momentos, y sin embargo la probabilidad de superar el valor 1,5 es de aproximadamente 0,067 en la primera y exactamente 0 en la segunda. Media y varianza no determinan la distribución; y las preguntas de la inferencia —"¿cuán raro sería observar una estimación tan alejada de cero como la que tengo delante, si en realidad β_2 fuese cero?"— son preguntas sobre probabilidades de colas, no sobre momentos.

Conviene señalar también que la lección 11 ya nos dio, sin que lo pareciera, la pieza que hoy resultará decisiva. Allí escribimos $\hat{\beta}_2 = \beta_2\, \mathit{1} + \sum_i W_i U_i$, y esa descomposición no era un paso intermedio prescindible: dice que el error de estimación $\hat{\beta}_2 - \beta_2$ es una *suma ponderada de las perturbaciones*. Si supiéramos qué forma tiene la distribución de las U_i, sabríamos qué forma tiene la de $\hat{\beta}_2$. Eso es lo que vamos a hacer: suponer una forma para U y leer la consecuencia.

Una última advertencia sobre el reparto de la sesión. Las cuatro primeras transparencias construyen el aparato; la quinta y la sexta lo traducen a geometría; las cuatro últimas se dedican a interpretarlo. Esa segunda mitad no es un apéndice ni un relleno: es, con bastante probabilidad, la parte de todo el curso que usted va a usar más veces —cada vez que lea un cuadro de resultados, propio o ajeno—.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, introducción y sección 4.1 (por qué la distribución muestral, y no solo sus momentos, es lo que hace falta para contrastar hipótesis).

2 El supuesto nuevo: normalidad

Condicionada a $\boldsymbol{X}$, cada perturbación U_i es *normal* de media 0 y varianza σ^2.

La descomposición de la lección 11 hace todo el trabajo:

$$\hat{\beta}_2 = \beta_2\, \mathit{1} + \sum_i W_i U_i.$$

Condicionando en $\boldsymbol{X}$ los pesos W_i son constantes, y una combinación lineal de normales independientes es normal:

$$\boxed{\hat{\beta}_2 \,\Big|\, \boldsymbol{X} \;\sim\; N\Big(\beta_2,\; \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}\Big).}$$

Lo *nuevo* es la palabra "N"; y es un supuesto genuino: el laboratorio anterior mostró que insesgadez y varianza no la necesitan.

Geométricamente: la normalidad hace el ruido *isótropo*, sin dirección privilegiada en $\mathbb{R}^n$.

El supuesto nuevo: normalidad

El Modelo Clásico de Regresión Lineal de la lección 9 constaba de tres supuestos —linealidad, independencia lineal de los regresores y homocedasticidad—, a los que la lección 11 añadió el

muestreo aleatorio. Hoy incorporamos un supuesto más, y conviene decir con precisión qué añade y qué no: la *normalidad* de la perturbación. Formalmente, suponemos que, condicionada a la matriz aleatoria de regresores $\boldsymbol{X}$, cada U_i tiene distribución normal:

$$U_i \,\Big|\, \boldsymbol{X} \;\sim\; N(0, \sigma^2), \qquad i = 1, \ldots, n.$$

Obsérvese que este supuesto no cambia ninguno de los momentos que ya teníamos: la media condicional sigue siendo $E[U_i \mid \boldsymbol{X}] = 0$ y la varianza condicional sigue siendo $Var[U_i \mid \boldsymbol{X}] = \sigma^2\, 1$, exactamente como en la lección 11. Lo único que añade es *de qué forma* se reparte la probabilidad alrededor de esa media, con esa varianza.[1]

De aquí a la distribución de $\hat{\beta}_2$ el camino es corto, y lo interesante es que no requiere ninguna cuenta nueva: basta releer lo que ya escribimos en la lección 11. Allí obtuvimos, por álgebra pura y sin tomar esperanzas,

$$\hat{\beta}_2 = \beta_2\, 1 + \sum_i W_i U_i, \qquad \text{con} \qquad W_i = \frac{X_i - \overline{X}}{\sum_j (X_j - \overline{X})^2}.$$

Los pesos W_i son funciones de $\boldsymbol{X}$, de modo que, una vez que condicionamos en $\boldsymbol{X}$, se comportan como números fijos —es exactamente el Lema que tanto usamos en la lección 11—. Y entonces $\sum_i W_i U_i$ es, condicionalmente, una combinación lineal de variables normales independientes (independientes por el muestreo aleatorio de la lección 11) con coeficientes constantes. Aquí usamos un resultado de su curso de probabilidad que no vamos a demostrar: *una combinación lineal de variables normales independientes es normal*. Con ello, condicionando en $\boldsymbol{X}$, la distribución de $\hat{\beta}_2$ es normal, y su media y su varianza son las que ya calculamos en la lección 11:

$$\hat{\beta}_2 \,\Big|\, \boldsymbol{X} \;\sim\; N\Big(\beta_2,\; \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}\Big).$$

Estrictamente hablando, esta expresión es una afirmación sobre la distribución *condicional*: dice que, fijada una realización concreta de $\boldsymbol{X}$, la distribución de $\hat{\beta}_2$ es la normal indicada, cuya varianza depende de esa realización (es, como vimos, una variable aleatoria si se la mira antes de observar $\boldsymbol{X}$). Es la misma disciplina de capas que venimos manteniendo desde la lección 9, y conviene no relajarla: la aleatoriedad que cuenta aquí es la de la perturbación, con los regresores dados.

Merece la pena insistir en por qué la normalidad es un supuesto *añadido* y no una consecuencia de los anteriores, porque el laboratorio que precede a esta lección lo puso delante de los ojos. Allí se repitió el experimento de Monte Carlo con perturbaciones uniformes y con perturbaciones χ^2 centradas —claramente no normales— y el resultado fue que la media de las estimaciones seguía cayendo sobre β_2 y su varianza empírica seguía coincidiendo con la fórmula de la lección 11. No podía ser de otro modo: aquella demostración nunca usó la normalidad. Lo que *sí* cambiaba era la forma del histograma. Y cambiaba de una manera instructiva: con $n = 506$ el histograma parecía

[1] *Una precisión notacional.* Dentro del símbolo $N(0, \sigma^2)$, los argumentos 0 y σ^2 son *parámetros* de una distribución —dos números—, no variables aleatorias. Por eso aquí escribimos 0 a secas, y no 0: la regla del curso que obliga a escribir $\beta_1\, 1$ en lugar de β_1 se aplica cuando un escalar se *combina* (se suma, se multiplica) con variables aleatorias, para mantener el paralelismo dimensional; no se aplica a los argumentos de una distribución. Compárese: $E[U_i \mid \boldsymbol{X}] = 0$ es una igualdad entre variables aleatorias (la esperanza condicional lo es), mientras que "la media de esa normal es 0" es una afirmación sobre un número.

normal aunque U no lo fuese, mientras que con $n = 12$ y una χ^2 de un grado de libertad la asimetría era inconfundible. Volveremos sobre esto al final de la sesión, porque es la razón por la que la inferencia que hoy construimos sigue siendo aproximadamente válida en muestras grandes incluso cuando la normalidad falla.

Queda la lectura geométrica, que es la que conecta este supuesto con el resto del curso. Pensemos en el vector aleatorio de perturbaciones $\boldsymbol{U} = (U_1, \ldots, U_n)$ como un "punto al azar" de $\mathbb{R}^n$. La homocedasticidad dice que todas sus componentes tienen la misma varianza; el muestreo aleatorio de la lección 11 dice que son *independientes* entre sí —no solo incorreladas, y aquí el matiz importa—. Las dos cosas juntas, *más* la normalidad de cada componente, dan algo más fuerte que la suma de sus partes: que la distribución de $\boldsymbol{U}$ es *esféricamente simétrica*, es decir, que no privilegia ninguna dirección de $\mathbb{R}^n$. Dicho de otro modo: la *dirección* en la que apunta el ruido es completamente uniforme, y su distribución no cambia si giramos los ejes. A esa propiedad la llamaremos *isotropía*, y es lo que hará que, en la sección "Por qué el valor crítico depende de $n-k$", tenga sentido hablar de "la distribución del ángulo". Sin normalidad, homocedasticidad e independencia siguen valiendo, pero la nube de ruido puede tener direcciones preferentes, y entonces el ángulo ya no tiene una distribución conocida.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.1 (supuesto MLR.6 de normalidad y distribuciones muestrales normales de los estimadores MCO).

3 El error estándar, y el precio de estimar σ

Esa varianza contiene σ^2, que *no conocemos*: la estimamos con la cuasivarianza de los residuos (lección 11), $\mathfrak{s}^2 = \mathrm{SRC}/(n-k)$.

Sustituyendo σ por $\mathfrak{s}$ sobre los datos observados, el *error estándar* de $\hat{\beta}_2$:

$$\boxed{\mathrm{ee}(\hat{\beta}_2) = \frac{\mathfrak{s}}{\sqrt{\sum_i (x_i - \mu_{\boldsymbol{x}})^2}} = \frac{\mathfrak{s}}{\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e}.}$$

Un cociente de longitudes: *ruido estimado por unidad de longitud del regresor en desviaciones.*

El precio de estimar σ: la normal se convierte en una *t de Student con $n-k$ grados de libertad* (enunciado, no demostrado):

$$\frac{\hat{\beta}_2 - \beta_2}{\mathrm{ee}(\hat{\beta}_2)} \;\sim\; t_{n-k}.$$

El error estándar, y el precio de estimar σ

La fórmula de la transparencia anterior tiene un problema práctico evidente: depende de σ^2, la varianza de la perturbación poblacional, que es una cantidad desconocida —como toda cantidad poblacional—. La lección 11 ya nos dejó resuelto ese punto: σ^2 se estima mediante la cuasivarianza de los residuos, $\mathfrak{s}^2 = \mathrm{SRC}/(n-k)$, que es el tercer miembro de la familia de productos escalares del curso, el de divisor $n-k$. Recordemos que ese divisor no era caprichoso: los residuos viven en el subespacio ortogonal a las columnas de $\mathbf{X}$, de dimensión $n-k$, y dividir por esa dimensión es lo que hace insesgado al estimador.

Sustituyendo σ por $\mathfrak{s}$ en la raíz cuadrada de la varianza, y escribiendo la expresión sobre los *datos observados* en lugar de sobre la muestra aleatoria, obtenemos el *error estándar* de $\hat{\beta}_2$:

$$\text{ee}(\hat{\beta}_2) = \frac{\mathfrak{s}}{\sqrt{\sum_i (x_i - \mu_{\boldsymbol{x}})^2}} = \frac{\mathfrak{s}}{\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e}.$$

La notación merece un comentario, porque el término "error estándar" es una de las fuentes de confusión más frecuentes al leer resultados. Un error estándar mide la dispersión de un *estimador* en su distribución muestral: cuánto variaría el número que hemos calculado si volviéramos a tomar otra muestra. No mide la dispersión de unos datos. Y el objeto que acabamos de escribir es, estrictamente, un *número* —se calcula con los datos que tenemos—, aunque sea la realización de una variable aleatoria (la que resultaría de evaluar esa misma fórmula sobre la muestra aleatoria). Es el mismo abuso de lenguaje, cómodo y universal, que ya señalamos en la lección 11 al hablar del sesgo de una fórmula escrita con datos.[2]

Escrito así, el error estándar es un *cociente de dos longitudes*, y esa es su lectura más útil. En el numerador, $\mathfrak{s}$ mide el tamaño típico del ruido. En el denominador, $\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e$ mide cuánto se separa el regresor de su propia media: es, literalmente, la longitud euclídea del vector en desviaciones de la lección 4. Por tanto: *el error estándar es el ruido estimado por unidad de longitud del regresor en desviaciones*. Un regresor con mucha variación —un vector en desviaciones largo— permite estimar su coeficiente con precisión; un regresor cuyos valores apenas cambian dentro de la muestra —un vector en desviaciones corto— no, por mucho que el ruido sea pequeño. Es la lectura geométrica de algo que en la lección 11 aparecía como "más dispersión en X, más precisión", y la veremos dibujada en la transparencia del intervalo de confianza.

Queda el segundo asunto de esta transparencia, que es sutil y conviene no pasar por alto. Si conociéramos σ, de la distribución normal de la transparencia anterior se seguiría inmediatamente que

$$\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\sigma^2 / \sum_i (X_i - \overline{X})^2}} \;\sim\; N(0,1),$$

porque restar la media y dividir por la desviación típica convierte cualquier normal en la normal estándar. Pero no conocemos σ: en su lugar usamos $\mathfrak{s}$, que es una variable aleatoria más, calculada con los mismos datos. Al dividir por una cantidad que también fluctúa de muestra en muestra, el cociente resultante se dispersa más que una normal estándar: puede salir grande porque el numerador sea grande, pero también porque el denominador haya salido pequeño por azar. El resultado exacto —que *enunciamos sin demostrar*, porque su prueba exige la distribución χ^2 y un argumento de independencia entre $\hat{\beta}_2$ y $\mathfrak{s}^2$ que este curso no desarrolla— es que ese cociente sigue una distribución conocida, la *t de Student con $n-k$ grados de libertad*:

$$\frac{\hat{\beta}_2 - \beta_2}{\text{ee}(\hat{\beta}_2)} \;\sim\; t_{n-k}.$$

[2] La confusión se agrava con la traducción. En inglés, la salida de un programa como Gretl distingue "S.D. dependent var" (*standard deviation*: dispersión de los datos del regresando) de "S.E. of regression" (*standard error*) y de la columna "Std. Error" de cada coeficiente. En español, las tres suelen aparecer rotuladas con alguna variante de "desviación típica", de modo que tres cantidades conceptualmente distintas comparten etiqueta. Volveremos sobre ello en la transparencia dedicada a leer una tabla de resultados.

La distribución t_{n-k} es simétrica y acampanada como la normal, pero con colas más gruesas, y tanto más gruesas cuanto menor sea $n-k$; a medida que $n-k$ crece, se acerca a la normal estándar hasta hacerse indistinguible de ella en la práctica (con $n-k$ del orden de unas decenas, la diferencia ya es pequeña). Que el número de grados de libertad sea exactamente $n-k$ —el mismo que aparecía en el divisor de $\mathfrak{s}^2$— no es casualidad, y en la sección "Por qué el valor crítico depende de $n-k$" veremos que tiene un significado geométrico muy concreto: es la dimensión del subespacio donde vive el residuo.

Nota de coherencia: en este curso, k cuenta *todos* los regresores, incluida la constante. Por eso los grados de libertad se escriben siempre $n-k$ y nunca $n-k-1$. En el modelo lineal simple de hoy, $k=2$ y los grados de libertad son $n-2$.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.2 (el estadístico t y su distribución); y cap. 3, sección 3.6 (errores estándar de los coeficientes).

4 El contraste t

Hipótesis nula: $H_0\colon \beta_2 = 0$ ("el regresor no aporta nada"), frente a $H_1\colon \beta_2 \neq 0$.

Estadístico (bajo H_0, sustituyendo $\beta_2 = 0$ en la transparencia anterior):

$$t = \frac{\hat{\beta}_2}{\mathrm{ee}(\hat{\beta}_2)} \;\sim\; t_{n-k}.$$

Decisión al nivel α (habitualmente 0,05): se rechaza H_0 si $|t| > t_c$, el valor crítico de la t_{n-k} que deja $\alpha/2$ en cada cola.

Valor p: probabilidad de un $|t|$ *al menos tan grande* como el obtenido, *si H_0 fuese cierta*. Se rechaza si es menor que α.

Nada obliga a contrastar contra cero: para $H_0\colon \beta_2 = c$ (p. ej. una elasticidad unitaria), $t = (\hat{\beta}_2 - c)/\mathrm{ee}(\hat{\beta}_2)$.

El contraste t

Con la distribución de la transparencia anterior en la mano, el contraste se construye solo. La hipótesis que interesa con más frecuencia es $H_0\colon \beta_2 = 0$, que en términos del modelo poblacional dice que el regresor no interviene: si $\beta_2 = 0$, entonces $Y = \beta_1\, 1 + U$ y X no aparece. La alternativa habitual es $H_1\colon \beta_2 \neq 0$, que no especifica ningún valor concreto y admite desviaciones en los dos sentidos (de ahí que el contraste se llame *de dos colas*).

El razonamiento es el de todo contraste de hipótesis, y conviene enunciarlo con cuidado porque es donde se cometen los errores de interpretación. *Suponemos provisionalmente* que H_0 es cierta. Bajo ese supuesto, sustituyendo $\beta_2 = 0$ en el resultado de la transparencia anterior, sabemos exactamente cómo se distribuye el cociente

$$t = \frac{\hat{\beta}_2}{\mathrm{ee}(\hat{\beta}_2)} \;\sim\; t_{n-k}.$$

Este cociente mide la distancia entre la estimación y el valor 0, pero medida en unidades de su propia imprecisión: dice cuántos errores estándar separan $\hat{\beta}_2$ de cero. Si el valor que obtenemos con nuestros datos es de los que la t_{n-k} produce con facilidad, los datos son compatibles con H_0 y no hay nada que objetar. Si es un valor que la t_{n-k} produciría muy raramente, tenemos un conflicto: o bien hemos tenido muy mala suerte, o bien el supuesto provisional era falso. El contraste opta por lo segundo, y al hacerlo acepta un riesgo cuantificado de equivocarse.

Ese riesgo es el *nivel de significación* α, la probabilidad de rechazar H_0 siendo cierta, que se fija de antemano (por convención, $\alpha = 0{,}05$; también se usan 0,10 y 0,01). Fijado α, el *valor crítico* t_c es el número que deja una probabilidad $\alpha/2$ en cada cola de la t_{n-k}, y la regla es: rechazar H_0 si $|t| > t_c$. El conjunto de valores que llevan a rechazar se llama *región crítica*. Nótese que t_c depende de dos cosas: del nivel α que hayamos elegido y de los grados de libertad $n - k$. Esta segunda dependencia es la que tiene una lectura geométrica sorprendente, y le dedicaremos una transparencia entera.

Una forma equivalente y, en la práctica, más informativa de presentar la misma decisión es el *valor p*: la probabilidad de que la t_{n-k} produzca un valor tan alejado de cero, o más, que el t que hemos obtenido —siempre, se insiste, bajo el supuesto de que H_0 es cierta—. Con él, la regla de decisión se vuelve inmediata: se rechaza H_0 si el valor p es menor que α. Su ventaja es que no obliga a comprometerse con un α antes de mirar los datos y permite al lector aplicar el suyo: un valor p de 0,048 y otro de 0,000001 llevan ambos a "rechazar al 5 %", pero no dicen lo mismo. Su inconveniente es que se malinterpreta con enorme facilidad, y a eso dedicaremos parte de la sección "Significación estadística $\neq$ relevancia económica".

Conviene, por último, deshacer una asociación demasiado automática entre "contraste t" y "contrastar contra cero". El cero tiene un lugar privilegiado porque corresponde a la pregunta "¿interviene este regresor?", pero la construcción vale para cualquier valor de referencia: para contrastar $H_0 \colon \beta_2 = c$ basta usar

$$t = \frac{\hat{\beta}_2 - c}{\mathrm{ee}(\hat{\beta}_2)},$$

que bajo esa H_0 sigue igualmente una t_{n-k}. Y hay preguntas económicas que exigen precisamente eso. Si el modelo está escrito en logaritmos y β_2 es una elasticidad (algo que veremos con detalle más adelante en el curso), la pregunta interesante rara vez es "¿es la elasticidad distinta de cero?" —casi siempre lo es— sino "¿es compatible con ser unitaria?", es decir, $H_0 : \beta_2 = 1$. Del mismo modo, si la teoría predice un valor concreto, contrastar contra ese valor es lo que pone a prueba la teoría; contrastar contra cero, no.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, secciones 4.2a-4.2e (contrastes de una y dos colas, valores críticos, valores p, y contrastes de hipótesis con valores de referencia distintos de cero).

5 Lectura geométrica: la razón de catetos

Con $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} = \hat{\beta}_2(\boldsymbol{x} - \overline{\boldsymbol{x}})$ y $\mathfrak{s}^2 = \mathrm{SRC}/(n-k)$, el estadístico queda:

$$\boxed{t^2 = (n-k)\,\frac{\mathrm{SEC}}{\mathrm{SRC}}, \qquad \text{es decir} \qquad |t| = \sqrt{n-k}\;\frac{\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_e}{\|\widehat{\boldsymbol{e}}\|_e}.}$$

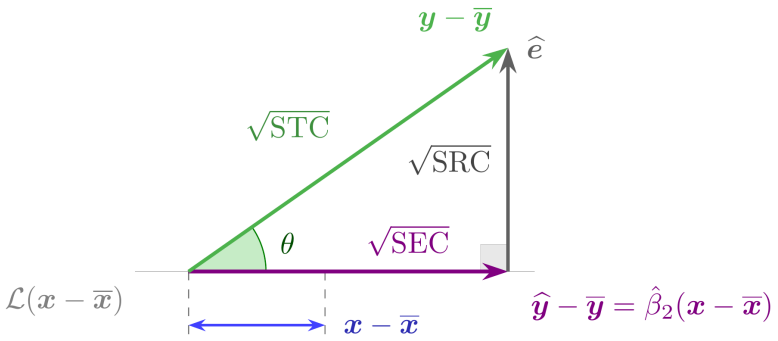


Figura 1: El triángulo rectángulo de la lección 8, con los dos catetos medidos. En verde, el vector de datos en desviaciones (la hipotenusa, de longitud $\sqrt{\text{STC}}$); en violeta, el ajuste en desviaciones (el cateto explicado, de longitud $\sqrt{\text{SEC}}$), rotulado a la derecha de su punta, que es un múltiplo del regresor en desviaciones (en azul, medido bajo el eje); en gris, el residuo (el cateto residual, de longitud $\sqrt{\text{SRC}}$). El ángulo θ es el mismo de $R^2 = \cos^2\theta$. El estadístico t compara los dos catetos.

Y como $\text{STC} = \text{SEC} + \text{SRC}$, también $t^2 = (n-k)R^2/(1-R^2)$: *el contraste t no mide nada que el R^2 no midiera ya.*

Lectura geométrica: la razón de catetos

Hasta aquí, la construcción del contraste ha sido puramente probabilística. Esta transparencia y la siguiente la traducen al lenguaje geométrico del curso, y el resultado es, creo, la mejor razón para haber hecho todo el recorrido anterior: el estadístico t resulta ser un objeto que ya conocíamos desde la lección 8, mirado con otra escala.

Empecemos por la pieza que hace posible la traducción. En el modelo lineal simple, el ajuste es $\widehat{\boldsymbol{y}} = \hat{\beta}_1\mathbf{1} + \hat{\beta}_2\boldsymbol{x}$, y sabemos desde la lección 7 que el ajuste y los datos tienen la misma media, $\mu_{\widehat{\boldsymbol{y}}} = \mu_{\boldsymbol{y}}$. Recordando además la notación de la lección 4 —la barra sobre un vector denota el vector constante formado por su media, $\overline{\boldsymbol{v}} = \mu_{\boldsymbol{v}}\mathbf{1}$—, y usando $\mu_{\widehat{\boldsymbol{y}}} = \hat{\beta}_1 + \hat{\beta}_2\mu_{\boldsymbol{x}}$, se obtiene

$$\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} = (\hat{\beta}_1\mathbf{1} + \hat{\beta}_2\boldsymbol{x}) - (\hat{\beta}_1 + \hat{\beta}_2\mu_{\boldsymbol{x}})\mathbf{1} = \hat{\beta}_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}) = \hat{\beta}_2(\boldsymbol{x} - \overline{\boldsymbol{x}}).$$

Es decir: el cateto explicado del triángulo de la lección 8 es el regresor en desviaciones multiplicado por $\hat{\beta}_2$. Esto es propio de la regresión simple —con dos o más regresores no constantes el ajuste en desviaciones ya no está alineado con ningún regresor individual, como advertimos en la lección 10—, y es la razón por la que hoy trabajamos con $k = 2$.

Tomando normas euclídeas al cuadrado en esa igualdad obtenemos $\text{SEC} = \hat{\beta}_2^2\,\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e^2$. A partir de aquí la cuenta es mecánica. Partimos del estadístico, escribimos el error estándar como el cociente de longitudes de la transparencia anterior y elevamos al cuadrado:

$$t = \frac{\hat{\beta}_2}{\text{ee}(\hat{\beta}_2)} = \frac{\hat{\beta}_2\,\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e}{\mathfrak{s}}, \qquad t^2 = \frac{\hat{\beta}_2^2\,\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e^2}{\mathfrak{s}^2} = \frac{\text{SEC}}{\text{SRC}/(n-k)} = (n-k)\,\frac{\text{SEC}}{\text{SRC}}.$$

que es la identidad anunciada. ■

Conviene detenerse en lo que dice esta igualdad, porque es fácil pasar por encima. A la izquierda, un estadístico de contraste, construido con supuestos probabilísticos, normalidad y distribuciones muestrales. A la derecha, dos sumas de cuadrados que calculamos en la lección 8 sin la menor referencia a la probabilidad, y un número entero que cuenta dimensiones. La única cantidad de la derecha que no viene de la pura geometría de los datos es $n-k$. Y la lectura es directa: el estadístico t compara la longitud del cateto explicado con la del cateto residual. Un cateto explicado largo frente a un residuo corto —es decir, un ángulo θ pequeño— da un $|t|$ grande; un cateto explicado corto frente a un residuo largo —un ángulo próximo a 90°— da un $|t|$ pequeño.

Nótese que la identidad está escrita para $|t|$, no para t: el cociente de longitudes es siempre positivo, mientras que t hereda el signo de $\hat{\beta}_2$. Si $\hat{\beta}_2$ es negativo, el cateto explicado apunta en el sentido opuesto sobre la misma recta $\mathcal{L}(\boldsymbol{x} - \overline{\boldsymbol{x}})$ (el conjunto de todos los múltiplos del regresor en desviaciones; geométricamente, una recta de $\mathbb{R}^n$), y nada más cambia: las longitudes, el ángulo y el valor de $|t|$ son los mismos. Por eso la figura está dibujada con $\hat{\beta}_2 > 0$ sin pérdida de generalidad.

La segunda forma de la identidad se obtiene recordando dos cosas de la lección 8: que $R^2 = \mathrm{SEC}/\mathrm{STC}$ y que $\mathrm{STC} = \mathrm{SEC} + \mathrm{SRC}$ (el teorema de Pitágoras aplicado al triángulo de la figura, en su versión con norma euclídea, sin el factor $1/n$ del producto escalar estadístico). De la segunda, $\mathrm{SRC} = \mathrm{STC} - \mathrm{SEC}$, y dividiendo numerador y denominador por STC:

$$\frac{\mathrm{SEC}}{\mathrm{SRC}} = \frac{\mathrm{SEC}/\mathrm{STC}}{1 - \mathrm{SEC}/\mathrm{STC}} = \frac{R^2}{1 - R^2}, \qquad \text{luego} \qquad t^2 = (n-k)\,\frac{R^2}{1-R^2}.$$

Y como $R^2 = \cos^2\theta$, donde $\theta = \measuredangle\big((\boldsymbol{y} - \overline{\boldsymbol{y}}), (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})\big)$, resulta que $|t|$ es una función únicamente del ángulo θ y del número $n-k$. No hay ninguna otra información en el estadístico t: dado el ángulo y dados los grados de libertad, $|t|$ está determinado.

Esto tiene una consecuencia práctica que conviene retener, y que reaparecerá en la sección "Significación estadística ≠ relevancia económica": un R^2 alto y un $|t|$ alto no son dos evidencias independientes que se refuercen mutuamente. Son la misma medida, expresada dos veces. Lo que $|t|$ añade al R^2 no es información sobre la relación, sino el factor $\sqrt{n-k}$: cuántas observaciones respaldan ese ángulo.

Podemos comprobar la identidad con el ejemplo numérico de la lección 8, donde $\mathrm{STC} = 100$, $\mathrm{SEC} = 90$, $\mathrm{SRC} = 10$ y $n = 5$ (con $k = 2$, luego $n - k = 3$):

$$t^2 = 3 \cdot \frac{90}{10} = 27, \qquad |t| = 5{,}20.$$

Y, por la otra vía, $R^2 = 0{,}9$ da $t^2 = 3 \cdot \frac{0{,}9}{0{,}1} = 27$, el mismo número. El valor crítico de la t_3 al 5 % es 3,18, de modo que aquel ajuste de juguete habría resultado significativo —con solo cinco observaciones—.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (donde aparece, para el contraste F, la misma relación entre estadísticos de contraste y sumas de cuadrados que aquí hemos obtenido para t).

6 Por qué el valor crítico depende de $n - k$

La identidad anterior separa la significación en *dos* factores: $|t| = \sqrt{n-k} \times \sqrt{\text{SEC}}/\sqrt{\text{SRC}}$, es decir, *cuánta evidencia* × *qué fuerza tiene la relación*.

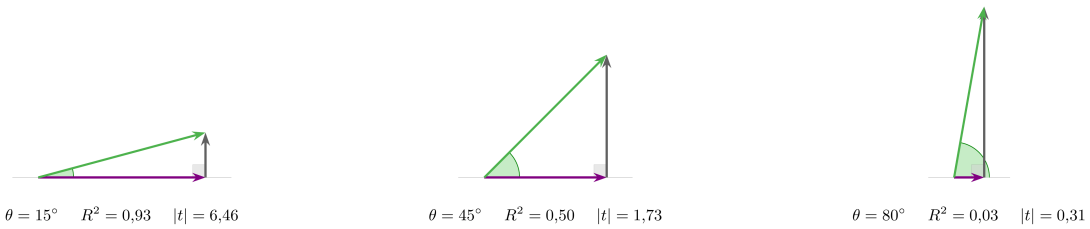


Figura 2: El mismo triángulo con tres aberturas del ángulo, y su R^2 y su $|t|$ (calculados con $n = 5$, es decir, $n-k = 3$, cuyo valor crítico al 5 % es 3,18). La hipotenusa mide lo mismo en los tres paneles: lo único que cambia es el ángulo. Solo el primero resulta significativo. La hipótesis $H_0 : \beta_2 = 0$ corresponde al caso límite en que el vector de datos en desviaciones es ortogonal a la dirección del regresor: ángulo de 90°, cateto explicado nulo.

Con pocas observaciones un ángulo pequeño no prueba nada: en $\mathbb{R}^2$ *todo* par de vectores en desviaciones está alineado ($R^2 = 1$ siempre, $n - k = 0$). En dimensión alta dos direcciones sin relación son *casi ortogonales*: el mismo ángulo sorprende más cuanto mayor es n.

Por qué el valor crítico depende de $n - k$

La identidad de la transparencia anterior admite una lectura que va más allá de lo anecdótico: descompone la significación estadística en sus dos ingredientes, y los separa limpiamente. Uno es geométrico y describe *la relación*: la razón entre los dos catetos, equivalentemente el ángulo θ, equivalentemente el R^2. El otro es aritmético y describe *la evidencia*: el factor $\sqrt{n-k}$, que solo cuenta observaciones. Un mismo ángulo, con más observaciones, produce un $|t|$ mayor; un mismo número de observaciones, con un ángulo más cerrado, también. La significación es el producto de ambas cosas, y por eso —como veremos— no puede leerse como si midiera solo una de las dos.

Detengámonos en la pregunta que da título a esta sección: ¿por qué hay que corregir por los grados de libertad? ¿Por qué un ángulo pequeño no es, por sí solo, evidencia suficiente? La respuesta más contundente está ya demostrada en este curso, y no en un rincón: es el resultado de la lección 5 según el cual *en* $\mathbb{R}^2$ *la correlación es siempre* ± 1. Recordemos el argumento: el subespacio $\mathcal{L}(\mathbf{1})^\perp$ —donde viven todos los vectores en desviaciones— tiene dimensión $n - 1$, de modo que en $\mathbb{R}^2$ es una simple recta; y dos vectores cualesquiera sobre una recta están forzosamente alineados.

Traduzcamos eso al lenguaje de hoy. Con $n = 2$ observaciones y $k = 2$ coeficientes, la recta de regresión pasa exactamente por los dos puntos: el residuo es nulo, SRC = 0, el ángulo θ es 0° y $R^2 = 1$. Un ajuste perfecto. ¿Es eso evidencia de una relación entre las variables? Evidentemente no: es evidencia de que dos puntos determinan una recta. Y el aparato que hemos construido lo detecta sin necesidad de que se lo advirtamos: los grados de libertad son $n - k = 2 - 2 = 0$, la

cuasivarianza $\mathfrak{s}^2 = \mathrm{SRC}/(n-k)$ es un cociente de la forma 0/0, el error estándar no está definido y el estadístico t no existe. No es que el contraste dé un resultado ambiguo: es que *no hay contraste posible*. Con $n = 3$ ya hay un grado de libertad, y el contraste existe, pero su valor crítico al 5 % es 12,71: exige una barbaridad de evidencia para rechazar, porque con tres puntos un ángulo pequeño sigue siendo fácil de obtener por azar.

Conviene leer la figura de la transparencia con esto en la mente. Los tres paneles tienen la misma hipotenusa y solo cambia el ángulo: con $\theta = 15°$ el R^2 es de 0,93 y $|t| = 6{,}46$, por encima del valor crítico 3,18; con $\theta = 45°$, un R^2 de 0,50 ya no basta ($|t| = 1{,}73$); y con $\theta = 80°$ el cateto explicado es casi nulo. El caso límite de ese tercer panel es precisamente la hipótesis nula: $\hat{\beta}_2 = 0$ equivale a que el vector de datos en desviaciones sea ortogonal a la dirección del regresor —ángulo de 90°, cateto explicado nulo, SEC = 0—.

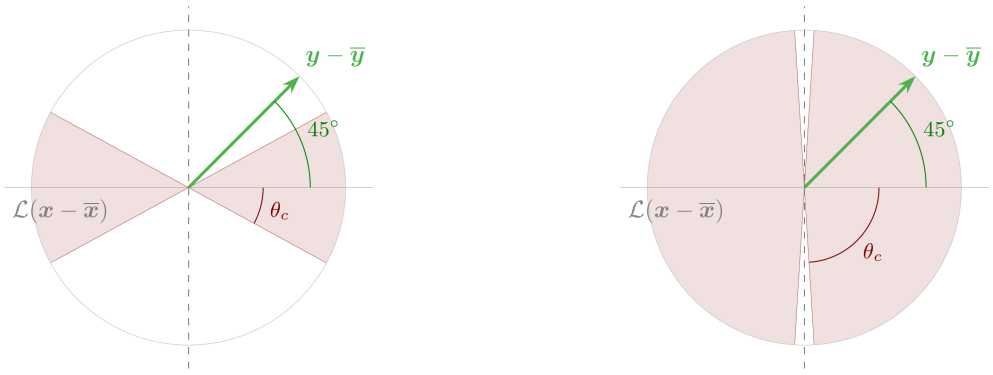
El fenómeno general es el recíproco de ese caso extremo, y es una propiedad de los espacios de dimensión alta que vale la pena enunciar aunque no la demostremos: *en $\mathbb{R}^n$ con n grande, dos direcciones tomadas sin relación entre sí son casi ortogonales*. Dicho de otra forma: si el vector de datos en desviaciones apunta "a un sitio cualquiera", lo esperable es que forme con la dirección del regresor un ángulo próximo a 90°, y tanto más próximo cuanto mayor sea n. Por eso un ángulo de, digamos, 45° es poco informativo con cinco observaciones —cabe fácilmente en el azar— y muy informativo con quinientas.

Aquí es donde el supuesto de normalidad de la sección "El supuesto nuevo: normalidad" cobra su sentido geométrico pleno, y donde podemos enunciar —sin demostrarlo— el resultado que cierra el círculo. Bajo $H_0\colon \beta_2 = 0$, el vector de datos en desviaciones es el vector de perturbaciones centrado; y si el ruido es isótropo, su *dirección* es uniforme: no privilegia ninguna orientación dentro de $\mathcal{L}(\mathbf{1})^{\perp}$. Entonces el ángulo θ entre esa dirección uniforme y la dirección fija del regresor tiene una distribución perfectamente determinada, que depende únicamente de una dimensión: la de lo que queda de $\mathcal{L}(\mathbf{1})^{\perp}$ una vez descontada la dirección del regresor, que es $n-k$ —el mismo subespacio donde vive el residuo—. Y como $|t|$ es una función del ángulo y de $n-k$ —eso es lo que dice la identidad de la transparencia anterior—, *la distribución t_{n-k} es, literalmente, la distribución del ángulo bajo H_0*. La normalidad no era, por tanto, un supuesto técnico arbitrario: es lo que hace que el ángulo tenga una distribución conocida.

Esto permite mirar la región crítica con otros ojos. Rechazar H_0 cuando $|t| > t_c$ equivale, por la identidad, a rechazarla cuando el ángulo es menor que un cierto ángulo crítico θ_c. Es decir: la región crítica es un *cono* alrededor de la recta generada por el regresor en desviaciones, y el valor crítico t_c no es más que la *abertura* de ese cono. Lo que hace el contraste es preguntar si el vector de datos en desviaciones ha caído dentro del cono.

La figura pone números a la idea. La abertura del cono se obtiene de la identidad: el ángulo crítico es aquel cuya razón de catetos vale $t_c/\sqrt{n-k}$. Con $n = 5$ ($n-k = 3$, $t_c = 3{,}18$) sale $\theta_c = 28{,}6°$: el cono es estrecho, y hace falta que los datos se alineen mucho con el regresor para rechazar. Con $n = 1002$ ($n-k = 1000$, $t_c = 1{,}96$) sale $\theta_c = 86{,}5°$: el cono ocupa casi todo el espacio, y basta un ángulo apenas distinto de la ortogonalidad para rechazar. Los dos paneles muestran el *mismo* vector de datos, a 45° del regresor: fuera del cono en el primer caso (no se rechaza, $|t| = 1{,}73$), dentro en el segundo (se rechaza, $|t| = 31{,}6$).

Hay que ser cuidadoso al interpretar ese contraste, porque admite dos lecturas y solo una es



$n = 5 \quad (n - k = 3), \quad t_c = 3{,}18$
abertura $\theta_c = 28{,}6°$
$|t| = 1{,}73$: **no se rechaza** H_0

$n = 1002 \quad (n - k = 1000), \quad t_c = 1{,}96$
abertura $\theta_c = 86{,}5°$
$|t| = 31{,}6$: **se rechaza** H_0

Figura 3: La región crítica (sombreada) como un cono alrededor de la recta generada por el regresor en desviaciones, vista dentro del plano de las desviaciones. A la izquierda, con $n = 5$: la abertura es de 28,6°. A la derecha, con $n = 1002$: la abertura es de 86,5°, casi todo el espacio. El vector de datos en desviaciones forma en los dos paneles el mismo ángulo de 45° con el regresor, y el veredicto es opuesto.

correcta. La lectura *incorrecta* sería: "con muchas observaciones el contraste es más laxo, rechaza con menos exigencia". No: el contraste mantiene, en los dos paneles, la misma probabilidad de error de tipo I, el 5 % —eso es lo que significa haber calculado la abertura con t_c—. Lo que cambia es que, con mil observaciones, el azar solo produce ángulos muy próximos a 90°: ese cono enorme es el 5 % de los casos porque casi toda la probabilidad se concentra en la rendija que queda fuera de él. La lectura correcta es, entonces: con muchas observaciones, una desviación pequeñísima de la ortogonalidad ya es estadísticamente detectable. Que sea *económicamente relevante* es otra cuestión, y a ella dedicaremos una transparencia entera.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.2b (dependencia del valor crítico respecto de los grados de libertad). Fisher, R. A. (1915), "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population", *Biometrika* 10, 507-521: el artículo donde la distribución del coeficiente de correlación se obtiene, originalmente, por un argumento geométrico sobre ángulos como el que aquí solo hemos enunciado.

7 El intervalo de confianza

Un *rango* de valores compatibles con los datos, en vez de un veredicto:

$$\boxed{\left[\ \hat{\beta}_2 - t_c \cdot \mathrm{ee}(\hat{\beta}_2)\ ,\quad \hat{\beta}_2 + t_c \cdot \mathrm{ee}(\hat{\beta}_2)\ \right]}$$

con el mismo t_c del contraste; semiamplitud $t_c\, \mathfrak{s}/\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e$.

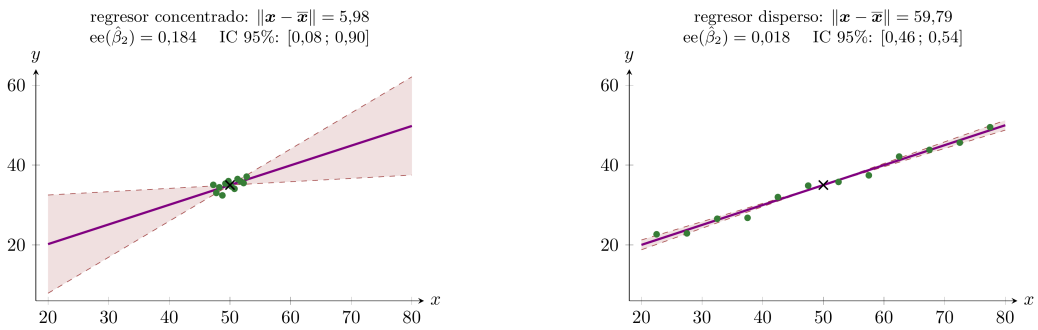


Figura 4: Dos ajustes con *la misma* muestra de perturbaciones, el mismo $n = 12$ y el mismo residuo —y por tanto el mismo $\mathfrak{s} = 1{,}10$—. Lo único distinto es cuánto se separan los valores del regresor. La zona sombreada es el haz de rectas cuyas pendientes caen en el intervalo de confianza al 95 %, dibujadas pivotando sobre el centroide, por donde la recta de regresión pasa siempre. Diez veces más largo el regresor en desviaciones, diez veces más estrecho el haz.

Dice: el 95 % de las muestras dan un intervalo que contiene a β_2. *No dice*: que β_2 esté en *este* con probabilidad 0,95; β_2 no es aleatorio, el intervalo sí.

El intervalo de confianza

El contraste de hipótesis responde a una pregunta muy pobre: sí o no. El intervalo de confianza responde a una mejor —"¿qué valores de β_2 son compatibles con estos datos?"— y con la misma maquinaria, sin ningún ingrediente adicional. Su construcción sale de despejar en la distribución de la sección "El error estándar, y el precio de estimar σ": si $(\hat{\beta}_2 - \beta_2)/\mathrm{ee}(\hat{\beta}_2)$ sigue una t_{n-k}, entonces ese cociente cae entre $-t_c$ y $+t_c$ con probabilidad $1-\alpha$, y despejando β_2 de esa doble desigualdad se obtiene

$$\Big[\ \hat{\beta}_2 - t_c \cdot \mathrm{ee}(\hat{\beta}_2)\ ,\quad \hat{\beta}_2 + t_c \cdot \mathrm{ee}(\hat{\beta}_2)\ \Big].$$

Es, por tanto, el mismo objeto que el contraste, presentado del revés. Y esa equivalencia es exacta y conviene tenerla clara: un valor c queda *fuera* del intervalo de confianza al 95 % si y solo si el contraste de $H_0\colon \beta_2 = c$ se rechaza al 5 %. El intervalo es, literalmente, el conjunto de las hipótesis que *no* se rechazarían. Por eso contiene toda la información del contraste y algo más: no solo dice si el cero se rechaza, sino qué otros valores también, y con qué margen.

La semiamplitud del intervalo, $t_c\,\mathfrak{s}/\|\boldsymbol{x}-\overline{\boldsymbol{x}}\|_e$, vuelve a tener la lectura geométrica de la sección "El error estándar, y el precio de estimar σ", ahora con consecuencias visibles. En la figura, esa semiamplitud es la abertura de la zona sombreada: el haz de todas las rectas cuya pendiente cae dentro del intervalo, dibujadas pivotando sobre el centroide $(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}})$, por el que la recta de regresión pasa siempre. Los dos paneles de la figura comparten absolutamente todo excepto la dispersión del regresor: los mismos parámetros ($\beta_1 = 10$, $\beta_2 = 0{,}5$), las mismas doce perturbaciones y —esto es lo que hace la comparación tan limpia— el mismo vector de residuos, porque el regresor en desviaciones apunta en la misma dirección en los dos casos y solo cambia su longitud.[3] En consecuencia $\mathfrak{s} = 1{,}10$

[3] En los dos paneles el regresor toma valores igualmente espaciados alrededor de 50; lo único que cambia es el paso

en ambos, y toda la diferencia procede del denominador: $\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_e$ vale 5,98 en el panel izquierdo y 59,79 en el derecho. Diez veces más largo el regresor en desviaciones, diez veces más pequeño el error estándar (0,184 frente a 0,018) y diez veces más estrecho el intervalo: [0,08; 0,90] frente a [0,46; 0,54].

Y no es solo el intervalo lo que se estrecha: el error de estimación efectivamente cometido también se divide por diez. La pendiente verdadera es $\beta_2 = 0{,}5$; con el regresor concentrado se estima 0,4931 (error de $-0{,}0069$) y con el regresor disperso, 0,4993 (error de $-0{,}00069$). Exactamente un factor de diez, el mismo que separa las longitudes. La lectura sustantiva es la que importa: *para estimar bien el efecto de una variable hace falta que esa variable varíe*. Si en su muestra todo el mundo tiene aproximadamente la misma educación, no podrá medir con precisión el efecto de la educación sobre el salario, por muy limpios que sean sus datos y por muchas observaciones que tenga. Esta idea reaparecerá, con otro disfraz, cuando estudiemos la colinealidad.

Vale la pena señalar que el panel izquierdo es un buen ejemplo de una situación incómoda pero frecuente: el contraste rechaza $H_0 : \beta_2 = 0$ (el cero no está en el intervalo, $|t| = 2{,}68$ frente a un crítico de 2,23), y sin embargo el intervalo es tan ancho —de 0,08 a 0,90— que apenas informa sobre la magnitud del efecto. Si un resultado así se comunica diciendo solo "el efecto es significativo", se está ocultando lo esencial. Es el argumento principal para preferir el intervalo al veredicto, sobre todo al leer o escribir un informe.

Queda la cuestión de la interpretación, que es donde casi todo el mundo resbala al menos una vez. La afirmación correcta es sobre el *procedimiento*, no sobre este intervalo concreto: si repitiéramos el muestreo muchas veces y en cada muestra construyéramos el intervalo con esta receta, el 95 % de esos intervalos contendrían al verdadero β_2. La afirmación incorrecta, y muy tentadora, es "hay un 95 % de probabilidad de que β_2 esté entre 0,46 y 0,54". No tiene sentido en este marco: β_2 es un número fijo y desconocido, no una variable aleatoria, así que no hay ninguna distribución de probabilidad sobre sus valores posibles. Lo aleatorio es el intervalo —sus dos extremos son variables aleatorias, funciones de la muestra—, no el parámetro. Este intervalo concreto, el que tenemos delante, o contiene a β_2 o no lo contiene; el 95 % describe con qué frecuencia acierta el método, no con qué probabilidad acierta esta vez.[4] La cobertura de los intervalos —qué fracción de ellos acierta realmente— es lo que comprobaremos con datos simulados en el laboratorio que sigue a esta lección.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.3 (intervalos de confianza y su relación con los contrastes de dos colas).

(0,5 frente a 5). Al ser el mismo patrón multiplicado por una constante, el vector en desviaciones tiene la misma dirección y una longitud diez veces mayor, y la proyección del ruido sobre esa dirección —y por tanto el residuo— es idéntica. De ahí que SRC y $\mathfrak{s}$ coincidan exactamente en ambos paneles, y que toda la diferencia recaiga sobre el denominador del error estándar.

[4] La distinción no es una pedantería: hay enfoques de la estadística —los llamados bayesianos— en los que sí tiene sentido asignar una distribución de probabilidad a un parámetro desconocido, y en los que existe un objeto (el *intervalo de credibilidad*) del que sí puede decirse "contiene a β_2 con probabilidad 0,95". Son objetos distintos, construidos con supuestos distintos, y a veces numéricamente parecidos. El de este curso es el intervalo de confianza clásico, y su garantía es la que se ha descrito: sobre el procedimiento.

8 Cómo se lee una tabla de resultados

Toda salida de estimación tiene, para cada regresor, las mismas cuatro columnas:

Columna	Qué es	A qué pregunta responde
Coeficiente	$\hat{\beta}_j$	¿de qué *tamaño* es el efecto, en las unidades del problema?
Desv. típica	$\mathrm{ee}(\hat{\beta}_j)$	¿con cuánta *precisión* lo hemos medido?
Estadístico t	el cociente de las dos anteriores	¿a cuántos errores estándar del cero está?
Valor p	$\mathrm{P}(\lvert t_{n-k}\rvert > \lvert t\rvert)$	¿cuán raro sería esto si el efecto fuese nulo?

Las dos últimas son *derivadas*: no añaden información, reordenan la de las dos primeras. La comprobación a mano —t = coeficiente / desviación típica— conviene hacerla siempre.

Aviso terminológico: en la salida en español, "desviación típica" rotula **tres** cantidades distintas: la de la variable dependiente (dispersión de los datos), la de la regresión ($\mathfrak{s}$) y la de cada coeficiente (su error estándar). En inglés solo la primera es *standard deviation*; las otras dos son *standard error*.

Cómo se lee una tabla de resultados

Todo lo construido hasta aquí se presenta, en la práctica, en forma de tabla: cuatro columnas por regresor, más unas pocas líneas de resumen del modelo. Conviene saber leerla despacio una vez para poder leerla rápido siempre. Tomemos como ejemplo el panel derecho de la figura de la transparencia anterior —el del regresor disperso—, cuya estimación completa es esta:

Regresor	Coeficiente	Desv. típica	Estadístico t	Valor p
constante	10,0343	0,9732	10,31	$1{,}2 \cdot 10^{-6}$
$\boldsymbol{x}$	0,4993	0,0184	27,14	$1{,}1 \cdot 10^{-10}$

con $n = 12$, $k = 2$, STC = 903,40, SEC = 891,30, SRC = 12,10, $\mathfrak{s} = 1{,}1001$ y $R^2 = 0{,}9866$.

La *primera columna* es la única que contiene información sustantiva nueva: el tamaño del efecto estimado, en las unidades del problema. Aquí, $\hat{\beta}_2 = 0{,}4993$ significa que a un aumento de una unidad de x le corresponde, en el ajuste, un aumento de 0,4993 unidades de y. Toda interpretación económica empieza —y muchas veces acaba— aquí. Es también la única columna cuyo valor cambia si cambian las unidades de medida: si midiéramos x en centenares, el coeficiente se multiplicaría por cien.

La *segunda columna* es el error estándar, y responde a una pregunta distinta: no cuánto vale el efecto, sino con cuánta precisión lo hemos medido. Es la columna que se ignora con más frecuencia y la que más falta hace: un coeficiente sin su error estándar es un número sin unidades de incertidumbre, y no se puede evaluar.

La *tercera y la cuarta columnas son derivadas*: se calculan a partir de las dos primeras y no añaden ninguna información que no estuviera ya en ellas. El estadístico t es el cociente de la primera entre la segunda, y el valor p es una transformación monótona de $|t|$ una vez fijados los grados de libertad. Esto tiene dos consecuencias prácticas. La primera es que conviene hacer la división a

mano de vez en cuando: 0,4993/0,0184 = 27,14, que es el t de la tabla. Es la comprobación más barata que existe contra un error de lectura de columna o contra una tabla mal transcrita en un artículo. La segunda es que, a diferencia del coeficiente, el estadístico t y el valor p *no* cambian si cambian las unidades de medida: el coeficiente y su error estándar se multiplican por el mismo factor y el cociente se conserva. Eso los hace cómodos para comparar y, al mismo tiempo, inservibles para valorar la magnitud del efecto.

Merece la pena leer también el pie de la tabla. El $R^2 = 0{,}9866$ dice, como sabemos desde la lección 8, que el ajuste explica el 98,66 % de la variación total de $\boldsymbol{y}$ —o, equivalentemente, que el ángulo θ es de unos 6,6°—. Y $\mathfrak{s} = 1{,}1001$ es la estimación de la desviación típica de la perturbación, expresada en las mismas unidades que $\boldsymbol{y}$: un dato útil, porque permite juzgar si el ruido es grande o pequeño *en términos del problema*, algo que el R^2, al ser un número adimensional, no dice. Podemos comprobar aquí la identidad de la sección "Lectura geométrica: la razón de catetos": $\sqrt{(n-k)\,\mathrm{SEC}/\mathrm{SRC}} = \sqrt{10 \cdot 891{,}30/12{,}10} = 27{,}14$, que es exactamente el $|t|$ de la tabla.

Queda el aviso terminológico, que en español es más necesario que en inglés. En la salida original de Gretl aparecen, en sitios distintos, "S.D. dependent var" (la desviación típica de la variable dependiente: una dispersión de *datos*), "S.E. of regression" (nuestra $\mathfrak{s}$: una estimación de la desviación típica de la *perturbación*) y la columna "Std. Error" de cada coeficiente (el error estándar propiamente dicho: la dispersión de un *estimador*). Son tres cantidades conceptualmente distintas, y la traducción española tiende a rotular las tres con alguna variante de "desviación típica", de modo que la distinción entre *desviación típica* y *error estándar* —que en inglés está en el propio nombre— se pierde. Al leer una salida en español conviene identificar cada una por su posición, no por su etiqueta.[5]

Una última observación sobre los artículos y los informes ajenos. Muchos no publican las cuatro columnas: es habitual dar el coeficiente y, debajo, entre paréntesis, el error estándar —o, en algunas tradiciones, el estadístico t—, con asteriscos que codifican niveles de significación. Antes de interpretar nada hay que averiguar qué contiene el paréntesis, porque la diferencia es enorme: un 0,02 bajo un coeficiente de 0,5 significa una estimación muy precisa si es un error estándar, y una estimación pésima si es un estadístico t. La nota al pie de la tabla lo suele decir; si no lo dice, la división a mano lo resuelve.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.6 (cómo se informan los resultados de una regresión, y cómo leer las convenciones habituales de presentación).

9 Significación estadística ≠ relevancia económica

La identidad de la sección "Lectura geométrica: la razón de catetos" lo explica sin necesidad de moraleja: $|t|$ mezcla *fuerza de la relación* con *cantidad de evidencia*.

[5] Hay, de hecho, una relación bonita entre las dos primeras. La "desviación típica de la variable dependiente" que imprimen los programas es $\sqrt{\mathrm{STC}/(n-1)}$, es decir, *exactamente* la misma fórmula que $\mathfrak{s} = \sqrt{\mathrm{SRC}/(n-k)}$ evaluada en el modelo con un único regresor, la constante ($k = 1$): en ese modelo el ajuste es el vector de medias y el residuo es el vector en desviaciones, luego SRC = STC y $n - k = n - 1$. La "dispersión de los datos" es, por tanto, el caso degenerado de la "dispersión residual" cuando el modelo no aporta nada más que la media.

n	R^2	θ	$\lvert t\rvert$	Veredicto al 5 %
3	0,90	18,4°	3,00	no significativo ($t_c = 12{,}71$)
5	0,90	18,4°	5,20	significativo ($t_c = 3{,}18$)
1002	0,01	84,3°	3,18	significativo ($t_c = 1{,}96$)

Mismo ángulo, veredicto opuesto (filas 1 y 2). Y un ángulo de 84° —casi ortogonal, R^2 del 1 %— resulta significativo con mil observaciones (fila 3).

Qué mirar en su lugar: el coeficiente en las unidades del problema; el intervalo de confianza; el R^2. La significación dice si el efecto es *distinguible de cero*, no si es *grande*.

Y tres avisos más: "no rechazar" no es "aceptar"; el valor p no es la probabilidad de que H_0 sea cierta; nada de esto establece causalidad.

Significación estadística ≠ relevancia económica

Esta es, si tuviera que elegir una, la transparencia más importante de la sesión, y la que tiene más probabilidades de resultarle útil fuera de este curso. Su contenido se resume en una frase: *"significativo" no quiere decir "grande"*. Y lo notable es que no hace falta apelar a la prudencia ni a la experiencia para justificarla: se lee directamente en la identidad que obtuvimos en la sección "Lectura geométrica: la razón de catetos".

Recordémosla: $|t| = \sqrt{n-k} \cdot \sqrt{\text{SEC}/\text{SRC}}$. El estadístico de contraste es el producto de dos factores que miden cosas completamente distintas. El segundo mide la *fuerza de la relación* en la muestra: el ángulo, el R^2, la razón de catetos. El primero solo cuenta *observaciones*. Un $|t|$ grande puede deberse a cualquiera de los dos, y el estadístico no distingue entre ellos: con suficientes observaciones, una relación arbitrariamente débil produce un $|t|$ arbitrariamente grande. Decir "el coeficiente es significativo" es, por tanto, decir algo sobre el producto de los dos factores, no sobre el efecto.

La tabla de la transparencia lleva esto al extremo con tres casos, calculados y no estimados a ojo. Las dos primeras filas tienen el *mismo* R^2 de 0,90 —el mismo ángulo de 18,4°, la misma fuerza de relación en la muestra— y veredictos opuestos: con $n = 3$ el estadístico vale 3,00 frente a un valor crítico de 12,71, y no se rechaza; con $n = 5$ vale 5,20 frente a 3,18, y se rechaza. La tercera fila hace el recorrido inverso: un R^2 del 1 % —un ángulo de 84,3°, prácticamente ortogonal, una relación casi inexistente— resulta significativo al 5 % en cuanto hay mil observaciones. Esta última situación no es una curiosidad de laboratorio: es lo que ocurre al trabajar con las bases de datos grandes que hoy son habituales, donde casi cualquier coeficiente resulta significativo y la significación deja, en consecuencia, de aportar información.

¿Qué mirar entonces? Tres cosas, por orden. *Primera*: el coeficiente, en las unidades del problema, y preguntarse si su magnitud es económicamente apreciable. Un efecto de dos céntimos sobre el salario mensual puede ser significativísimo y no interesarle a nadie; un efecto de doscientos euros puede no ser significativo y merecer toda la atención —y una muestra mayor—. Esta lectura del tamaño del coeficiente es un hilo que recorrerá el resto del curso: al introducir modelos con logaritmos y variables dicotómicas, buena parte del trabajo consistirá precisamente en saber qué significa $\hat{\beta}_j$ en cada caso (un cambio absoluto, un cambio porcentual, una elasticidad, un desplazamiento). *Segunda*: el intervalo de confianza, que dice de una vez si el efecto es distinguible de cero y con qué margen; el panel izquierdo de la figura de la transparencia anterior era un caso claro de resultado

significativo pero prácticamente informativo de nada. *Tercera*: el R^2, con toda la cautela que ya le pusimos en la lección 8 —no es una medida de calidad del modelo, y un R^2 alto puede ser puramente espurio—, pero que al menos informa de la fuerza de la relación sin mezclarla con el tamaño de la muestra.

Los tres avisos del último fragmento merecen desarrollo, porque son los errores de lectura más frecuentes.

"No rechazar" no es "aceptar". Cuando el contraste no rechaza $H_0 : \beta_2 = 0$, lo correcto es decir que los datos son compatibles con que β_2 sea cero; no que β_2 sea cero. La diferencia es la que hay entre "no he encontrado pruebas" y "he probado que no hay nada". Un contraste no rechaza por dos motivos muy distintos: porque el efecto es realmente nulo, o porque no hay evidencia suficiente para detectarlo —regresor poco disperso, muestra pequeña, mucho ruido—. El intervalo de confianza permite distinguir los dos casos: un intervalo estrecho alrededor del cero es informativo ("si hay efecto, es pequeño"); un intervalo ancho que contiene al cero no dice casi nada. Por eso, ante un resultado no significativo, la pregunta útil no es "¿acepto H_0?" sino "¿qué anchura tiene mi intervalo?".

El valor p no es la probabilidad de que H_0 sea cierta. Es la probabilidad de observar unos datos al menos tan extremos como los observados *suponiendo que H_0 es cierta*. Las dos cosas se parecen en la redacción y no tienen nada que ver: la primera exigiría una distribución de probabilidad sobre las hipótesis, que en este marco no existe —β_2 es un número fijo, como ya discutimos con el intervalo—. Un valor p de 0,03 no significa "hay un 3 % de probabilidad de que el regresor no importe". Significa: "si el regresor no importara, datos como estos aparecerían el 3 % de las veces".

Nada de esto establece causalidad. El contraste t se pronuncia sobre un coeficiente de un ajuste; de si ese coeficiente mide un efecto causal no dice, ni puede decir, absolutamente nada. Es la misma advertencia de la lección 9 —la descomposición ortogonal no implica causalidad— y de la lección 8 —las correlaciones espurias cumplen la identidad de Pitágoras con la misma fuerza matemática que las relaciones sensatas—. Un coeficiente significativo en una relación espuria es igual de significativo. La significación mide la compatibilidad de los datos con una hipótesis sobre un parámetro; la causalidad es una cuestión ajena, que exige teoría y, sobre todo, un diseño que permita sostenerla.

Conviene añadir un cuarto aviso, que apunta ya a las dos lecciones siguientes. Los contrastes t que hemos construido son *individuales*: cada uno se pronuncia sobre un coeficiente, por separado. Y en regresión múltiple puede ocurrir que ningún coeficiente sea individualmente significativo y que, sin embargo, los regresores en conjunto sí expliquen el regresando: eso es lo que detecta el contraste F de la próxima lección. Suele ser síntoma de que dos o más regresores apuntan en direcciones muy próximas y el ajuste no puede repartir el efecto entre ellos con precisión —la colinealidad, a la que dedicaremos otra lección—. Dicho en el lenguaje de hoy: no es que falte relación, es que a cada regresor le queda poca longitud *propia* con la que medirse.

Y, para cerrar, una observación sobre el supuesto de normalidad con el que abrimos la sesión. En el laboratorio anterior vimos que, con $n = 506$, el histograma de las estimaciones parecía normal aunque la perturbación no lo fuera, mientras que con $n = 12$ la asimetría era inconfundible. Eso no era casualidad: el teorema central del límite garantiza que, al crecer la muestra, la distribución de $\hat{\beta}_2$ se aproxima a la normal aunque U no sea normal. La consecuencia práctica es tranquilizadora y

conviene conocerla: en muestras grandes, los contrastes t y los intervalos de confianza de esta lección siguen siendo aproximadamente válidos sin el supuesto de normalidad. En muestras pequeñas, en cambio, ese supuesto es el que sostiene el resultado *exacto*, y si falla, el valor p que imprime el programa puede estar bastante equivocado.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.2f (significación económica frente a significación estadística: el epígrafe que más conviene leer de todo el capítulo); y cap. 5, sección 5.2 (normalidad asintótica de MCO y validez aproximada de los contrastes en muestras grandes).

10 Recapitulación y guiño a las sesiones siguientes

Pieza	Fórmula	Lectura geométrica
Error estándar	$\mathfrak{s}/\|\boldsymbol{x}-\overline{\boldsymbol{x}}\|_e$	ruido por unidad de longitud del regresor
Estadístico t	$\hat{\beta}_2/\mathrm{ee}(\hat{\beta}_2)$	$\sqrt{n-k}\times$ razón de catetos
Grados de libertad	$n-k$	dimensión del subespacio donde vive el residuo
Región crítica	$\|t\|>t_c$	cono alrededor de la dirección del regresor
Intervalo de confianza	$\hat{\beta}_2 \pm t_c \cdot \mathrm{ee}(\hat{\beta}_2)$	haz de pendientes compatibles

Un solo supuesto nuevo (normalidad); todo lo demás venía de las lecciones 8 y 11.

Lección 13: el mismo cociente de catetos, con dos divisores en lugar de uno, es el contraste F —que juzga varios regresores a la vez—. En regresión simple, $F = t^2$.

Más adelante: con k regresores, $\mathrm{ee}(\hat{\beta}_j) = \mathfrak{s}/\|\tilde{\boldsymbol{x}}_j\|_e$, donde $\tilde{\boldsymbol{x}}_j$ es lo que le queda de *propio* al regresor j tras descontar lo que comparte con los demás. Cuando dos regresores casi se solapan, esa longitud se acorta: colinealidad.

Recapitulación y guiño a las sesiones siguientes

Visto en conjunto, el balance de esta sesión es parecido al de la lección 11: mucho resultado con muy poco ingrediente nuevo. El único supuesto añadido ha sido la normalidad de la perturbación, y su único cometido ha sido dar *forma* a una distribución cuyos dos primeros momentos ya conocíamos. Todo lo demás —el triángulo, las sumas de cuadrados, el R^2, la varianza de $\hat{\beta}_2$, la cuasivarianza de los residuos con su divisor $n-k$— estaba ya sobre la mesa desde las lecciones 8 y 11.

La tabla resume las cinco piezas y su lectura geométrica. Conviene subrayar el patrón que las recorre: todas ellas son, en el fondo, comparaciones de longitudes dentro del plano de las desviaciones. El error estándar compara el ruido con la longitud del regresor; el estadístico t compara los dos catetos del triángulo; el intervalo de confianza convierte esa comparación en un abanico de pendientes; los grados de libertad cuentan las dimensiones disponibles para el residuo; y la región crítica es un cono cuya abertura fija el valor crítico. No hay, en toda la inferencia de esta lección, ningún objeto que no pueda mirarse así.

Dos hilos quedan abiertos, y los dos salen de la misma identidad.

El primero apunta a la lección siguiente. El cociente SEC/SRC que hoy ha aparecido dentro de t^2 no es propiedad del contraste t: es el cociente que compara la parte explicada con la parte no explicada, y admite una versión más general en la que cada una se divide por sus propios grados de libertad. Ese cociente —SEC/$(k-1)$ frente a SRC/$(n-k)$— es el *estadístico* F, que permite juzgar varios regresores a la vez en lugar de uno por uno. En el caso de hoy, con un único regresor no constante, $k-1=1$ y el F coincide con t^2: el mismo objeto con otro nombre. La figura del triángulo servirá, sin cambiar un trazo, para la próxima lección.

El segundo apunta algo más lejos. Todo lo de hoy se ha demostrado para $k=2$, con un único regresor no constante, porque solo en ese caso el cateto explicado es un múltiplo del regresor. Con más regresores, la fórmula del error estándar sigue teniendo la misma estructura —un cociente de longitudes—, pero el denominador cambia de una manera muy instructiva: en lugar de la longitud del regresor en desviaciones, aparece la longitud de $\tilde{\boldsymbol{x}}_j$, la parte del regresor j que *no* se puede reconstruir a partir de los demás regresores. Lo enunciamos aquí sin demostrarlo:

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

La lectura es inmediata y anticipa un problema que trataremos con detalle: si el regresor j es casi una combinación lineal de los demás —si apunta casi en la misma dirección que ellos—, entonces $\tilde{\boldsymbol{x}}_j$ es muy corto, el error estándar se dispara y el coeficiente se estima con muy poca precisión. No porque falte información en los datos, sino porque esa información no permite separar el efecto de ese regresor del de sus compañeros. Es la *colinealidad*, y es la traducción exacta, en el lenguaje de los ángulos de este curso, de lo que en la transparencia del intervalo de confianza vimos con un solo regresor: para medir bien el efecto de una variable hace falta que esa variable varíe *por su cuenta*.

Para profundizar: Wooldridge, J. M. (2020), cap. 4, sección 4.5 (el contraste F, con el que arranca la lección siguiente); y cap. 3, sección 3.4 (varianza de los estimadores MCO en regresión múltiple y el papel de la colinealidad).