Índice general

Lección 7. Estimación MCO sin matrices: dos ortogonalidades, dos coeficientes

Marcos Bujosa

19 de septiembre de 2026

Resumen

Hoy resolvemos el sistema de dos ortogonalidades planteado en la lección anterior: de él surgen las fórmulas de $\hat{\beta}_2$ y $\hat{\beta}_1$. La pendiente resulta ser la covarianza dividida por la varianza —o, equivalentemente, la correlación reescalada—. Por ahora sin matrices de por medio.

"Imponer perpendicularidad, dos veces, y despejar: así nace la recta de mínimos cuadrados."

- (slides) — (html) — (pdf)

1 De dónde venimos: un plano, un problema sin resolver

En la lección 6 planteamos la regresión simple como una proyección sobre el *plano* de todas las combinaciones lineales de $\mathbf{1}$ y $\boldsymbol{x}$, es decir, sobre $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$:

$$\boldsymbol{y} \quad = \quad \hat{\beta}_1 \mathbf{1} + \hat{\beta}_2 \boldsymbol{x} + \widehat{\boldsymbol{e}} \quad = \quad \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}.$$

- $\widehat{\boldsymbol{e}}$ debe ser ortogonal a *todo* el plano.
- Eso equivale a ser ortogonal a cada uno de sus generadores:

$$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0, \qquad \langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0.$$

En la lección 6 nos quedamos ahí: dos condiciones, ningún coeficiente despejado.

Hoy resolvemos ese sistema. Sin matrices: dos ecuaciones, dos incógnitas ($\hat{\beta}_1$, $\hat{\beta}_2$).

De dónde venimos: un plano, un problema sin resolver

Conviene recordar el camino recorrido antes de avanzar. En la lección 4 resolvimos el problema más sencillo posible de proyección: buscar, dentro de la recta $\mathcal{L}(\mathbf{1})$, el vector más cercano a $\boldsymbol{y}$. La condición de ortogonalidad del residuo[1] frente al único generador $\mathbf{1}$ nos dio, sin apenas esfuerzo, la media $\mu_{\boldsymbol{y}}$. En la lección 6 dimos el siguiente paso: añadimos un segundo generador, $\boldsymbol{x}$, y pasamos de una recta a un *plano*, $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$. La lógica no cambió en absoluto —seguíamos buscando el vector

[1] que en aquel caso denominamos vector en desviaciones

del subespacio más cercano a $\boldsymbol{y}$—, pero ahora la condición de ortogonalidad del residuo frente a *todo el plano* se traduce en *dos* condiciones, una por cada generador:

$$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0, \qquad \langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0.$$

La lección 6 dejó planteado este sistema —explicando por qué debía cumplirse— pero no lo resolvió. Esa es precisamente la tarea de hoy: determinar $\hat{\beta}_1$ y $\hat{\beta}_2$ a partir de esas dos ecuaciones, sin recurrir aún a ninguna maquinaria matricial (que llegará en la lección 10, y solo como notación compacta para expresar esto mismo).

Merece la pena insistir en que no hay ningún concepto nuevo en esta sesión. Todo lo que necesitamos —el producto escalar, la ortogonalidad, la media, la varianza, la covarianza— ya fue introducido y demostrado en las lecciones 2 a 5. Hoy simplemente *aplicamos* esas herramientas a un sistema de dos ecuaciones. Es, en cierto sentido, la culminación aritmética de todo lo anterior: la recompensa de haber invertido tiempo en entender qué es realmente un producto escalar.

También conviene recordar la motivación última de todo esto, sembrada ya en la lección 3: MCO no es más que la búsqueda del vector $\widehat{\boldsymbol{y}}$ del subespacio $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ que minimiza la distancia $\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|$. Esa minimización —un problema de mínimos cuadrados— es equivalente a imponer que el vector de error $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$ sea ortogonal al subespacio. Hoy vemos, por fin, adónde lleva esa condición cuando se traduce en números.

Para profundizar:

- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 2, sección 2.2: "Deriving the Ordinary Least Squares Estimates" (la misma derivación algebraica, en notación de sumatorios).

2 Primera condición: el residuo tiene media cero

Recordemos: $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \hat{\beta}_1 \mathbf{1} - \hat{\beta}_2 \boldsymbol{x}$.

De la primera condición, $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$:

$$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = \Big\langle (\boldsymbol{y} - \hat{\beta}_1 \mathbf{1} - \hat{\beta}_2 \boldsymbol{x}), \mathbf{1} \Big\rangle_s = \mu_{\boldsymbol{y}} - \hat{\beta}_1 - \hat{\beta}_2 \mu_{\boldsymbol{x}} = 0.$$

Despejando

$$\boxed{\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}.}$$

El intercepto (la constante) es *función de* la pendiente (aún falta una ecuación más).

Primera condición: el residuo tiene media cero

La primera condición de ortogonalidad, $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ ya la hemos visto en otro contexto: es la misma condición que, en la lección 4, hacía que el vector en desviaciones tuviera suma cero. Aquí el residuo es más complejo —depende de dos coeficientes, no de uno—, pero la condición algebraica es idéntica: la media de los residuos es cero, por tanto, *la suma de los residuos es cero.*

Recordando que: $\widehat{\boldsymbol{e}} = \boldsymbol{y} - (\hat{\beta}_1 \mathbf{1} + \hat{\beta}_2 \boldsymbol{x})$, la primera condición queda en:

$$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = \Big\langle (\boldsymbol{y} - \hat{\beta}_1 \mathbf{1} - \hat{\beta}_2 \boldsymbol{x}) \, , \, \mathbf{1} \Big\rangle_s = 0;$$

y aplicando la propiedad distributiva:

$$\langle \boldsymbol{y}, \mathbf{1} \rangle_s - \hat{\beta}_1 \langle \mathbf{1}, \mathbf{1} \rangle_s - \hat{\beta}_2 \langle \boldsymbol{x}, \mathbf{1} \rangle_s = 0.$$

Recordando la definición de media de un vector: $\langle \boldsymbol{v}, \mathbf{1} \rangle_s = \mu_{\boldsymbol{v}}$, y que $\langle \mathbf{1}, \mathbf{1} \rangle_s = \|\mathbf{1}\|_s^2 = 1$, llegamos a

$$\mu_{\boldsymbol{y}} - \hat{\beta}_1 - \hat{\beta}_2 \mu_{\boldsymbol{x}} = 0,$$

de donde despejamos inmediatamente

$$\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}.$$

Este resultado, todavía incompleto (pues depende del valor —aún desconocido— de $\hat{\beta}_2$), revelará el papel que juega $\hat{\beta}_1$ en el ajuste de la recta de regresión: sea cual sea la pendiente $\hat{\beta}_2$, la constante $\hat{\beta}_1$ siempre tomará el valor necesario para que la recta ajustada pase por el punto $(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}})$. Volveremos sobre esta idea, con su lectura geométrica completa, más adelante en la sesión.

Hay otra lectura de este mismo resultado, igual de importante. Recordemos que en la lección 6 vimos que $\mu_{\widehat{e}} = \langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ (el vector de residuos tiene media cero). Esto permite deducir que la ortogonalidad del residuo frente a $\mathbf{1}$ implica que el ajuste *preserva la media* de $\boldsymbol{y}$. En efecto: como $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$, tomando medias en ambos lados (la media es lineal, pues es un producto escalar) obtenemos

$$\mu_{\boldsymbol{y}} \;=\; \mu_{\widehat{\boldsymbol{y}}} + \mu_{\widehat{e}} \;=\; \mu_{\widehat{\boldsymbol{y}}}.$$

Es decir: *la media del vector ajustado coincide con la media del vector original*. Este hecho, que puede parecer una curiosidad menor, será la clave geométrica de la transparencia dedicada al lugar de las medias, más adelante en esta misma sesión.

Para profundizar:

- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: linealidad del producto escalar, aplicada aquí a la media de una suma de vectores.

3 Segunda condición: el residuo ortogonal a $\boldsymbol{x}$

De la segunda condición, $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$:

$$\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s \;=\; \Big\langle (\boldsymbol{y} - \hat{\beta}_1 \mathbf{1} - \hat{\beta}_2 \boldsymbol{x}), \boldsymbol{x} \Big\rangle_s \;=\; \langle \boldsymbol{y}, \boldsymbol{x} \rangle_s - \hat{\beta}_1 \, \mu_{\boldsymbol{x}} - \hat{\beta}_2 \, \mu_{\boldsymbol{x}^2} \;=\; 0.$$

Sustituyendo $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}$ (transparencia anterior) y agrupando:

$$\langle \boldsymbol{y}, \boldsymbol{x} \rangle_s - \mu_{\boldsymbol{x}} \mu_{\boldsymbol{y}} \;=\; \hat{\beta}_2 (\mu_{\boldsymbol{x}^2} - \mu_{\boldsymbol{x}}^2).$$

Como $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_s = \mu_{\boldsymbol{x} \odot \boldsymbol{y}}$ (media del producto componente a componente):

$$\underbrace{\mu_{\boldsymbol{x} \odot \boldsymbol{y}} - \mu_{\boldsymbol{x}} \mu_{\boldsymbol{y}}}_{\sigma_{\boldsymbol{xy}}} \;=\; \hat{\beta}_2 \big(\underbrace{\mu_{\boldsymbol{x}^2} - \mu_{\boldsymbol{x}}^2}_{\sigma_{\boldsymbol{x}}^2} \big);$$

donde aparecen las "fórmulas de cálculo" de covarianza y varianza. Así:

$$\boxed{\sigma_{\boldsymbol{xy}} = \hat{\beta}_2 \, \sigma_{\boldsymbol{x}}^2.}$$

Segunda condición: el residuo ortogonal a $\boldsymbol{x}$

La segunda condición de ortogonalidad, $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x}\rangle_s = 0$, se desarrolla con la misma herramienta que la primera —la propiedad distributiva (por la bilinealidad) del producto escalar—, aplicada ahora sobre $\boldsymbol{x}$ en lugar de sobre $\mathbf{1}$:

$$\langle \widehat{\boldsymbol{e}}, \boldsymbol{x}\rangle_s = \Big\langle (\boldsymbol{y} - \hat{\beta}_1\mathbf{1} - \hat{\beta}_2\boldsymbol{x})\,,\, \boldsymbol{x}\Big\rangle_s = \langle \boldsymbol{y}, \boldsymbol{x}\rangle_s - \hat{\beta}_1\langle \mathbf{1}, \boldsymbol{x}\rangle_s - \hat{\beta}_2\langle \boldsymbol{x}, \boldsymbol{x}\rangle_s = 0.$$

Necesitamos identificar los dos productos escalares que involucran a $\boldsymbol{x}$ consigo mismo o con $\mathbf{1}$. El primero ya lo conocemos: $\langle \mathbf{1}, \boldsymbol{x}\rangle_s = \mu_{\boldsymbol{x}}$ (la media, lección 4). El segundo lo vimos en la lección 5: $\langle \boldsymbol{x}, \boldsymbol{x}\rangle_s = \frac{1}{n}\sum_{i=1}^n x_i x_i = \mu_{\boldsymbol{x}^2}$, donde $\boldsymbol{x}^2$ denotaba el vector cuyas componentes son los cuadrados de las de $\boldsymbol{x}$ —aunque allí no lo dijimos con este nombre, $\boldsymbol{x}^2$ es $\boldsymbol{x} \odot \boldsymbol{x}$, el *producto componente a componente* de $\boldsymbol{x}$ consigo mismo—. Del mismo modo tenemos que $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_s = \frac{1}{n}\sum_{i=1}^n x_i y_i = \mu_{\boldsymbol{x}\odot\boldsymbol{y}}$, donde $\boldsymbol{x} \odot \boldsymbol{y}$ denota el vector cuyas componentes son los productos componente a componente de $\boldsymbol{x}$ e $\boldsymbol{y}$. Sustituyendo productos escalares tenemos:

$$\mu_{\boldsymbol{x}\odot\boldsymbol{y}} - \hat{\beta}_1\,\mu_{\boldsymbol{x}} - \hat{\beta}_2\,\mu_{\boldsymbol{x}^2} = 0.$$

Sustituimos ahora $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2\mu_{\boldsymbol{x}}$ (transparencia anterior) y agrupamos los términos con $\hat{\beta}_2$:

$$\mu_{\boldsymbol{x}\odot\boldsymbol{y}} - (\mu_{\boldsymbol{y}} - \hat{\beta}_2\mu_{\boldsymbol{x}})\mu_{\boldsymbol{x}} - \hat{\beta}_2\mu_{\boldsymbol{x}^2} = 0 \implies \mu_{\boldsymbol{x}\odot\boldsymbol{y}} - \mu_{\boldsymbol{x}}\mu_{\boldsymbol{y}} = \hat{\beta}_2(\mu_{\boldsymbol{x}^2} - \mu_{\boldsymbol{x}}^2).$$

El paréntesis de la derecha ya lo reconocemos: es exactamente $\sigma_{\boldsymbol{x}}^2 = \mu_{\boldsymbol{x}^2} - \mu_{\boldsymbol{x}}^2$, la "fórmula de cálculo" de la varianza que obtuvimos en la lección 5 a partir de Pitágoras. Así que

$$\mu_{\boldsymbol{x}\odot\boldsymbol{y}} - \mu_{\boldsymbol{x}}\mu_{\boldsymbol{y}} = \hat{\beta}_2\,\sigma_{\boldsymbol{x}}^2.$$

Falta reconocer el lado izquierdo; y resulta ser la "fórmula de cálculo" de la *covarianza* —el análogo exacto de la fórmula de la varianza, ahora con dos vectores distintos en lugar de uno consigo mismo—. En efecto, por la misma bilinealidad que hemos usado en toda esta lección (y que ya usamos en la lección 2):

$$\sigma_{\boldsymbol{x}\boldsymbol{y}} = \langle \boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1},\, \boldsymbol{y} - \mu_{\boldsymbol{y}}\mathbf{1}\rangle_s = \langle \boldsymbol{x}, \boldsymbol{y}\rangle_s - \mu_{\boldsymbol{x}}\langle \mathbf{1}, \boldsymbol{y}\rangle_s - \mu_{\boldsymbol{y}}\langle \boldsymbol{x}, \mathbf{1}\rangle_s + \mu_{\boldsymbol{x}}\mu_{\boldsymbol{y}}\langle \mathbf{1}, \mathbf{1}\rangle_s = \mu_{\boldsymbol{x}\odot\boldsymbol{y}} - \mu_{\boldsymbol{x}}\mu_{\boldsymbol{y}}.$$

Así que nuestra ecuación se reescribe, sin ningún ingrediente nuevo, como

$$\sigma_{\boldsymbol{x}\boldsymbol{y}} = \hat{\beta}_2\,\sigma_{\boldsymbol{x}}^2,$$

que es ya, prácticamente, el resultado final. Solo falta despejar $\hat{\beta}_2$, lo que hacemos en la siguiente transparencia.

Nota de notación. El subíndice de $\sigma_{\boldsymbol{x}\boldsymbol{y}}$ (covarianza *entre* $\boldsymbol{x}$ e $\boldsymbol{y}$, vectores ya centrados) y el de $\mu_{\boldsymbol{x}\odot\boldsymbol{y}}$ (media *del vector* $\boldsymbol{x}\odot\boldsymbol{y}$, sin centrar) se parecen tipográficamente, pero indican cosas distintas: el primero nombra una relación entre dos vectores; el segundo nombra la media de un vector nuevo, construido multiplicando componente a componente. El símbolo $\odot$ hace explícita esa construcción para evitar la confusión entre ambos subíndices.

Para profundizar:

- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: bilinealidad del producto escalar.

4 $\hat{\beta}_2$: covarianza sobre varianza

Despejando de $\sigma_{xy} = \hat{\beta}_2\, \sigma_x^2$:

$$\boxed{\hat{\beta}_2 \;=\; \frac{\sigma_{xy}}{\sigma_x^2}\,.}$$

Y recordando $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$ (lección 5):

$$\hat{\beta}_2 \;=\; \rho_{xy} \frac{\sigma_y}{\sigma_x}.$$

La pendiente MCO es *la correlación, reescalada por el ratio de las desviaciones típicas*: su signo es el de ρ_{xy}.

$\hat{\beta}_2$: covarianza sobre varianza

De la ecuación $\sigma_{xy} = \hat{\beta}_2 \sigma_x^2$ obtenida en la transparencia anterior se despeja de inmediato

$$\hat{\beta}_2 = \frac{\sigma_{xy}}{\sigma_x^2}.$$

Esta es la fórmula que probablemente recuerde de un curso de estadística: la pendiente de la recta de mínimos cuadrados es la covarianza entre $\boldsymbol{x}$ e $\boldsymbol{y}$ dividida por la varianza de $\boldsymbol{x}$. Lo que hemos hecho aquí es mostrar *de dónde sale*: no es una regla mnemotécnica aislada, sino la consecuencia directa —vía Pitágoras y bilinealidad del producto escalar— de imponer la segunda condición de ortogonalidad.

Podemos reescribir esta fórmula de una manera todavía más reveladora. Recordando la definición de correlación de la lección 5, $\rho_{xy} = \sigma_{xy}/(\sigma_x \sigma_y)$, despejamos $\sigma_{xy} = \rho_{xy}\, \sigma_x \sigma_y$ y sustituimos:

$$\hat{\beta}_2 = \frac{\rho_{xy}\, \sigma_x \sigma_y}{\sigma_x^2} = \rho_{xy} \frac{\sigma_y}{\sigma_x}.$$

Esta identidad, ya anticipada como promesa en las lecciones 5 y 6 ("pendiente = correlación reescalada"), tiene una lectura muy intuitiva: la pendiente MCO es el coseno del ángulo θ entre los vectores en desviaciones de $\boldsymbol{x}$ e $\boldsymbol{y}$ (la correlación), corregido por el cociente de las escalas de ambas variables (σ_y/σ_x). Si $\boldsymbol{x}$ e $\boldsymbol{y}$ tuvieran la misma dispersión ($\sigma_x = \sigma_y$), la pendiente coincidiría exactamente con la correlación.

Una consecuencia inmediata que conviene subrayar: como σ_x^2 y σ_y son siempre no negativas (son normas, o normas al cuadrado), el *signo* de $\hat{\beta}_2$ coincide siempre con el signo de σ_{xy}, que a su vez coincide siempre con el signo de ρ_{xy}. Dicho de otro modo: *la pendiente de la recta de regresión y la correlación entre las dos variables tienen siempre el mismo signo*. Esta observación, trivial una vez vista la fórmula, es la base de la intuición habitual de que "pendiente positiva" y "correlación positiva" son, en cierto sentido, la misma cosa vista desde ángulos distintos.

Para profundizar:

- Bujosa, M. *Curso de Álgebra Lineal*, libro online, cap. 11.
- Wooldridge, J. M. (2020), cap. 2, ecuación (2.19) y comentario adyacente sobre el signo de $\hat{\beta}_2$.

5 Vía alternativa: generadores ortogonales

Recordemos (lección 6): $\mathcal{L}(\mathbf{1}, \boldsymbol{x}) = \mathcal{L}(\mathbf{1},\ \boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$.

Y ahora sí: $\mathbf{1} \perp (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$, porque $\Big\langle (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})\,,\, \mathbf{1} \Big\rangle_s = \mu_{\boldsymbol{x}} - \mu_{\boldsymbol{x}} = 0$.

Con generadores *ortogonales*, la proyección se descompone en dos componentes ortogonales, uno por generador; cada coeficiente es la razón que resulta de proyectar sobre esa dirección *por separado* (Cauchy–Schwarz, lección 3: $\alpha = \frac{\langle \boldsymbol{a},\boldsymbol{b}\rangle}{\langle \boldsymbol{a},\boldsymbol{a}\rangle} = \frac{\langle \boldsymbol{a},\boldsymbol{b}\rangle}{\|\boldsymbol{a}\|^2}$):

$$c_1 = \frac{\langle \boldsymbol{y}\,,\,\mathbf{1}\rangle_s}{\|\mathbf{1}\|_s^2} = \mu_{\boldsymbol{y}}, \qquad c_2 = \frac{\Big\langle \boldsymbol{y}, (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}) \Big\rangle_s}{\|\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}\|_s^2} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}^2} = \hat{\beta}_2.$$

Sin resolver ningún sistema: dos proyecciones independientes, sobre dos direcciones perpendiculares.

Vía alternativa: generadores ortogonales

Existe un segundo camino hacia el mismo resultado, más elegante que resolver un sistema de ecuaciones, y que conviene conocer porque reaparecerá (en su versión con más de un regresor) en la lección 10.

La observación de partida ya se hizo en la lección 6: el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ —el conjunto de todas las combinaciones lineales de $\mathbf{1}$ y $\boldsymbol{x}$— puede generarse igualmente con $\mathbf{1}$ y el vector en desviaciones $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$:

$$\mathcal{L}(\mathbf{1}, \boldsymbol{x}) = \mathcal{L}(\mathbf{1},\ \boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}).$$

(Esto es así porque $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$ es una combinación lineal de $\mathbf{1}$ y $\boldsymbol{x}$, y viceversa: $\boldsymbol{x} = (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}) + \mu_{\boldsymbol{x}}\mathbf{1}$; cambiar de generadores de este modo no cambia el conjunto de combinaciones lineales que se pueden formar, es decir, no cambia el plano.)

La razón por la que este cambio de generadores es interesante —y no un mero capricho notacional— es que, a diferencia de $\mathbf{1}$ y $\boldsymbol{x}$ (que en general *no* son ortogonales, salvo que $\mu_{\boldsymbol{x}} = 0$, pues $\langle \mathbf{1}, \boldsymbol{x}\rangle_s = \mu_{\boldsymbol{x}}$), los nuevos generadores $\mathbf{1}$ y $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$ SÍ son ortogonales:

$$\Big\langle (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})\,,\, \mathbf{1} \Big\rangle_s = \langle \boldsymbol{x}, \mathbf{1}\rangle_s - \mu_{\boldsymbol{x}}\langle \mathbf{1}, \mathbf{1}\rangle_s = \mu_{\boldsymbol{x}} - \mu_{\boldsymbol{x}} \cdot 1 = 0.$$

¿Por qué importa la ortogonalidad de los generadores? Porque cuando un plano se genera con dos vectores ortogonales, la proyección de cualquier vector sobre ese plano se descompone en *dos proyecciones independientes*, una sobre cada generador —exactamente la fórmula de proyección de la lección 3, $\alpha = \frac{\langle \boldsymbol{b},\boldsymbol{a}\rangle}{\langle \boldsymbol{a},\boldsymbol{a}\rangle} = \frac{\langle \boldsymbol{b},\boldsymbol{a}\rangle}{\|\boldsymbol{a}\|^2}$ (la misma que sirvió para demostrar Cauchy–Schwarz), aplicada dos veces, una por generador. No hace falta resolver ningún sistema de ecuaciones simultáneas: cada coeficiente se calcula de forma completamente independiente del otro.

Aplicándolo aquí: sea $\widehat{\boldsymbol{y}} = c_1\mathbf{1} + c_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$ la proyección de $\boldsymbol{y}$ sobre el plano, escrita en esta nueva base. Por ortogonalidad de los generadores,

$$c_1 = \frac{\langle \boldsymbol{y}\,,\,\mathbf{1}\rangle_s}{\|\mathbf{1}\|_s^2} = \frac{\mu_{\boldsymbol{y}}}{1} = \mu_{\boldsymbol{y}}, \qquad c_2 = \frac{\Big\langle \boldsymbol{y}, (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}) \Big\rangle_s}{\|\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}\|_s^2} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}^2} = \hat{\beta}_2.$$

El mismo resultado que obtuvimos resolviendo el sistema, pero sin resolver ningún sistema.

Falta un último paso: traducir c_1, c_2 (los coeficientes en la base "centrada") a $\hat{\beta}_1, \hat{\beta}_2$ (los coeficientes en la base original $\mathbf{1}, \boldsymbol{x}$). Basta sustituir c_1 y c_2 en $\widehat{\boldsymbol{y}} = c_1\mathbf{1} + c_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$ y reagrupar:

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

Reconocemos, en el término entre paréntesis, exactamente la fórmula de $\hat{\beta}_1$ que ya habíamos obtenido de la primera condición de ortogonalidad. Los dos caminos —resolver el sistema directamente, o cambiar a generadores ortogonales— llegan, como no podía ser de otro modo, al mismo resultado. El segundo camino es bastante ilustrativo: muestra con claridad que la dificultad de resolver "un sistema de dos ecuaciones" era, en el fondo, solo un problema de *elección de base*; con la base adecuada (generadores ortogonales), el sistema se desacopla por completo.

Para profundizar:

- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: proyección sobre un vector, fórmula $\alpha = \langle \boldsymbol{a}, \boldsymbol{b}\rangle / \langle \boldsymbol{a}, \boldsymbol{a}\rangle$.
- Strang, G. (2016). *Introduction to Linear Algebra*, sección sobre bases ortogonales y proyecciones independientes.

6 $\hat{\beta}_1$ y el lugar de las medias

De $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2\mu_{\boldsymbol{x}}$, reordenando:

$$\mu_{\boldsymbol{y}} = \hat{\beta}_1 + \hat{\beta}_2\mu_{\boldsymbol{x}}.$$

Así, evaluando la recta ajustada $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ en el punto $x = \mu_{\boldsymbol{x}}$, el resultado es $\mu_{\boldsymbol{y}}$

La recta ajustada pasa exactamente por el punto de las medias $(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}})$.

Además (primera condición): $\mu_{\widehat{\boldsymbol{y}}} = \mu_{\boldsymbol{y}}$, el ajuste preserva la media.

Tres cantidades que coinciden: $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}} = \hat{\beta}_1 + \hat{\beta}_2\mu_{\boldsymbol{x}}$ (versión interactiva).

Y hay más, los vectores en desviaciones de $\widehat{\boldsymbol{y}}$ y $\hat{\beta}_2\boldsymbol{x}$ son *iguales*:

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

De ahí que $\sigma_{\widehat{\boldsymbol{y}}} = \sigma_{\hat{\beta}_2\boldsymbol{x}}$. Y también que $\boldsymbol{y}$ forme *el mismo ángulo* con $\widehat{\boldsymbol{y}}$ que con $\hat{\beta}_2\boldsymbol{x}$.

$\hat{\beta}_1$ y el lugar de las medias

La fórmula $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2\mu_{\boldsymbol{x}}$, obtenida ya en la transparencia "Primera condición: el residuo tiene media cero", admite una lectura geométrica muy bonita si la reordenamos como

$$\mu_{\boldsymbol{y}} = \hat{\beta}_1 + \hat{\beta}_2\mu_{\boldsymbol{x}}.$$

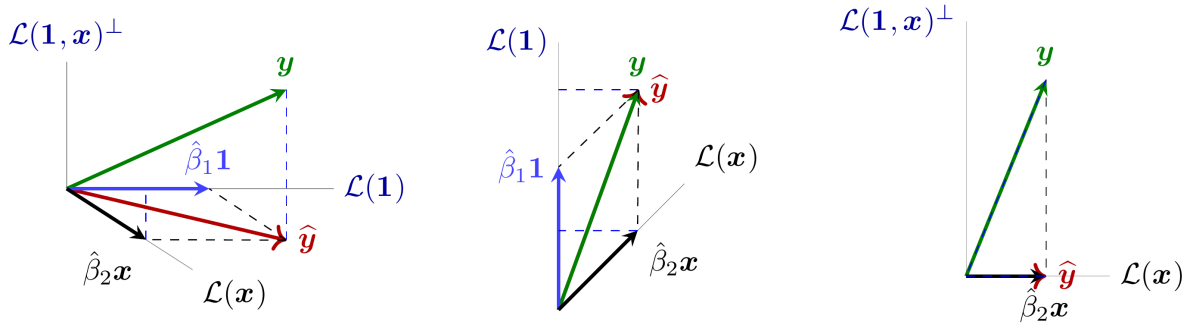

Figura 1: La proyección de la lección 6, vista desde tres ángulos, con las medias señaladas. (Izquierda) De frente al plano $(\mathcal{L}(\mathbf{1}), \mathcal{L}(\mathbf{1},\boldsymbol{x})^\perp)$: las proyecciones de $\boldsymbol{y}$ y de $\widehat{\boldsymbol{y}}$ sobre $\mathcal{L}(\mathbf{1})$ caen en el mismo punto. (Centro) Vista cenital: sumar la proyección de $\hat{\beta}_2\boldsymbol{x}$ y la de $\hat{\beta}_1\mathbf{1}$ sobre $\mathcal{L}(\mathbf{1})$ reconstruye ese mismo punto. (Derecha) Vista alineada con $\mathcal{L}(\mathbf{1})$: lo que se ve son, literalmente, los vectores en desviaciones; el ángulo entre el de $\boldsymbol{y}$ y el de $\widehat{\boldsymbol{y}}$ será protagonista de la lección 8.

Esta expresión dice, literalmente, que si evaluamos la recta ajustada $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ en el punto $x = \mu_{\boldsymbol{x}}$, el resultado es exactamente $\mu_{\boldsymbol{y}}$. Dicho de otro modo: *la recta de regresión MCO pasa siempre, sin excepción, por el punto de las medias* $(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}})$. Este resultado, que en un curso de estadística aparece a menudo como una propiedad más entre otras, es aquí una consecuencia directa e inevitable de la primera condición de ortogonalidad: no puede dejar de cumplirse.

Esta propiedad se conecta directamente con la observación que hicimos en la transparencia "Primera condición: el residuo tiene media cero": como $\widehat{\boldsymbol{e}} \perp \mathbf{1}$, se cumple $\mu_{\widehat{\boldsymbol{e}}} = 0$, y por tanto $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$: la media del vector ajustado coincide con la media del vector original. Podemos ahora reunir ambas piezas en una única cadena de igualdades:

$$\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}} = \hat{\beta}_1 + \hat{\beta}_2 \mu_{\boldsymbol{x}}.$$

La primera igualdad es un hecho sobre *vectores* (dos vectores distintos, $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$, comparten la misma media). La segunda es un hecho sobre la *recta ajustada*, evaluada en un punto concreto. Ambas son, en realidad, la misma consecuencia geométrica vista desde dos ángulos.

La figura recién mostrada más arriba hace visible, con tres perspectivas complementarias, toda esta cadena de igualdades. La vista de la izquierda retoma el plano frontal $(\mathcal{L}(\mathbf{1}), \mathcal{L}(\mathbf{1},\boldsymbol{x})^\perp)$ ya usado en la lección 6 (la vista central de aquella figura): en ella se aprecia directamente que $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$ tienen la misma proyección sobre $\mathcal{L}(\mathbf{1})$, es decir, $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$. La vista central es una vista *cenital*, perpendicular al plano de los regresores: en ella se ve que sumar la proyección de $\hat{\beta}_2\boldsymbol{x}$ (sobre $\mathcal{L}(\mathbf{1})$) a $\hat{\beta}_1\mathbf{1}$ reconstruye exactamente el punto $\mu_{\boldsymbol{y}}\mathbf{1}$; y que las longitudes de las componentes perpendiculares a $\mathcal{L}(\mathbf{1})$ de $\widehat{\boldsymbol{y}}$ y de $\hat{\beta}_2\boldsymbol{x}$ son iguales (es decir, que $\widehat{\boldsymbol{y}}$ y $\hat{\beta}_2\boldsymbol{x}$ tienen la misma desviación típica). La vista de la derecha, nueva, está tomada *alineada con* $\mathcal{L}(\mathbf{1})$. Al mirar en la dirección de $\mathbf{1}$ perdemos toda profundidad en esa dirección, y lo que vemos —de frente, sin distorsión— es el propio plano $\mathcal{L}(\mathbf{1})^\perp$. Es decir: vemos los vectores en desviaciones de $\boldsymbol{y}$, de $\widehat{\boldsymbol{y}}$ y de $\hat{\beta}_2\boldsymbol{x}$, con su verdadera longitud y su verdadero ángulo entre sí.

Esta última vista no es un simple adorno: adelanta, sin demostrarlo aún, al protagonista de la lección 8. El ángulo que en ella se aprecia entre $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$ —medido, como siempre, entre sus vectores en desviaciones— será exactamente el ángulo cuyo coseno al cuadrado define el R^2.

La igualdad que anuncian las vistas central y derecha —que las componentes perpendiculares de $\widehat{\boldsymbol{y}}$ y de $\hat{\beta}_2\boldsymbol{x}$ coinciden— no es una casualidad geométrica ni requiere ningún argumento nuevo: es consecuencia inmediata de una propiedad ya demostrada en la lección 4. Como $\widehat{\boldsymbol{y}} = \hat{\beta}_1\mathbf{1} + \hat{\beta}_2\boldsymbol{x}$ y $\hat{\beta}_1\mathbf{1}$ es un vector constante, sumarlo no cambia el vector en desviaciones (lección 4: "sumar una constante no cambia la desviación típica"; aquí usamos, más precisamente, que ni siquiera cambia el propio vector en desviaciones, no solo su norma). Por tanto:

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

El vector en desviaciones de $\widehat{\boldsymbol{y}}$ es, *literalmente*, el mismo vector que el vector en desviaciones de $\hat{\beta}_2\boldsymbol{x}$ (no solo un vector de la misma longitud: es el mismísimo vector). De ahí se siguen, sin ningún paso adicional, dos consecuencias:

- Tomando normas: $\sigma_{\widehat{\boldsymbol{y}}} = \sigma_{\hat{\beta}_2\boldsymbol{x}}$ (dos vectores idénticos tienen, evidentemente, la misma norma).
- Como la correlación de $\boldsymbol{y}$ con cualquier vector depende únicamente del vector en desviaciones de ese vector (lección 5), y $\widehat{\boldsymbol{y}}$ y $\hat{\beta}_2\boldsymbol{x}$ comparten el mismo vector en desviaciones, $\boldsymbol{y}$ forma el mismo ángulo con ambos.

Guardamos esta identidad —$\widehat{\boldsymbol{y}} - \mu_{\widehat{\boldsymbol{y}}}\mathbf{1} = \hat{\beta}_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$— porque reaparecerá en la lección 8: es la pieza que permite traducir directamente entre la descomposición de varianzas $\mathrm{STC} = \mathrm{SEC} + \mathrm{SRC}$ y la relación $R^2 = \rho^2_{\boldsymbol{xy}}$ de la regresión simple.

Conviene subrayar que esta propiedad —la recta pasa por el punto de las medias— es exclusiva del criterio de mínimos cuadrados *con constante incluida*. Si forzáramos, por alguna razón, un ajuste sin intercepto (es decir, proyectando solo sobre $\mathcal{L}(\boldsymbol{x})$, sin $\mathbf{1}$), esta propiedad desaparecería, porque ya no tendríamos la primera condición de ortogonalidad ($\langle\widehat{\boldsymbol{e}}, \mathbf{1}\rangle_s = 0$) que la garantiza. En este curso, todos los modelos de regresión incluyen constante, así que esta propiedad se cumple siempre. Es más: en opinión del autor de estos apuntes, **sin constante no puede hablarse de regresión**, porque nos salimos del marco de la estadística; no incluirla es un error conceptual. El laboratorio que sigue a la lección 8 muestra qué se pierde al omitirla.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, comentario tras la ecuación (2.18) sobre el punto de las medias.

7 Ejemplo numérico completo

Como en la actividad 3 (html) del laboratorio ($\boldsymbol{y} = 2 + 3\boldsymbol{x} + \boldsymbol{u}$), pero con $n = 5$, calculable a mano:

x_i	u_i	$y_i = 2 + 3x_i + u_i$	$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$
1	1	6	5
2	-2	6	8
3	0	11	11
4	2	16	14
5	-1	16	17

$$\mu_{\boldsymbol{x}} = 3, \quad \mu_{\boldsymbol{y}} = 11, \quad \sigma^2_{\boldsymbol{x}} = 2, \quad \sigma_{\boldsymbol{xy}} = 6.$$

$$\hat{\beta}_2 = \frac{6}{2} = 3, \qquad \hat{\beta}_1 = 11 - 3 \cdot 3 = 2.$$

¡Recuperamos exactamente $\beta_1 = 2$, $\beta_2 = 3$! (No es lo habitual: $\boldsymbol{u}$ se eligió ortogonal a los regresores.)

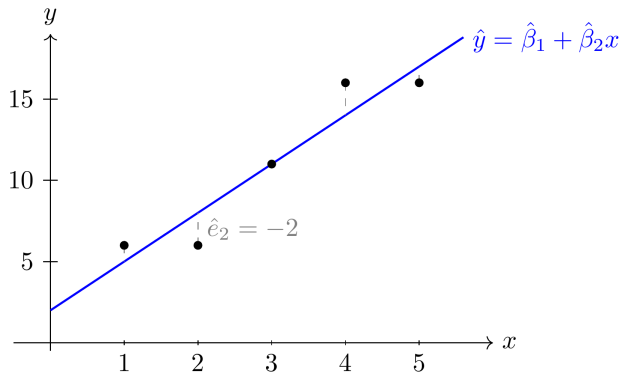

Figura 2: Diagrama de dispersión clásico del ejemplo: los cinco puntos (x_i, y_i) y la recta ajustada $\hat{y} = 2 + 3x$. Las líneas discontinuas son los residuos $\hat{e}_i = u_i$.

Versión interactiva: cuaderno 2 en MyBinder, parte 3 (qué le exigimos a $\boldsymbol{u}$ para recuperar los coeficientes).

Ejemplo numérico completo

Verifiquemos con números todo lo dicho hasta ahora. Construimos un ejemplo pequeño, calculable enteramente a mano, en el mismo espíritu que la actividad de la práctica de laboratorio (actividad 3 (html)), donde se generaban datos a partir de una relación lineal conocida $\boldsymbol{y} = 2 + 3\boldsymbol{x} + \boldsymbol{u}$ para comprobar que MCO recupera aproximadamente esos coeficientes generadores. Aquí, con solo $n = 5$ observaciones (frente a la muestra mayor y aleatoria de la práctica), podemos seguir el cálculo paso a paso.

Tomamos $\boldsymbol{x} = (1, 2, 3, 4, 5)$ y un vector $\boldsymbol{u} = (1, -2, 0, 2, -1)$ (nótese que $\sum u_i = 0$ y que, como veremos, $\boldsymbol{u}$ se eligió deliberadamente también ortogonal a $\boldsymbol{x}$). Generamos $y_i = 2 + 3x_i + u_i$:

$$\boldsymbol{y} = (6,\ 6,\ 11,\ 16,\ 16).$$

Medias: $\mu_{\boldsymbol{x}} = \frac{1+2+3+4+5}{5} = 3$, $\mu_{\boldsymbol{y}} = \frac{6+6+11+16+16}{5} = \frac{55}{5} = 11$.

Vectores en desviaciones: $\boldsymbol{x} - \overline{\boldsymbol{x}} = (-2,-1,0,1,2)$, $\boldsymbol{y} - \overline{\boldsymbol{y}} = (-5,-5,0,5,5)$.

Varianza de $\boldsymbol{x}$:

$$\sigma_{\boldsymbol{x}}^2 = \frac{1}{5}\left[(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2\right] = \frac{1}{5}(4+1+0+1+4) = \frac{10}{5} = 2.$$

Covarianza:

$$\sigma_{\boldsymbol{xy}} = \frac{1}{5}\left[(-2)(-5) + (-1)(-5) + (0)(0) + (1)(5) + (2)(5)\right] = \frac{1}{5}(10+5+0+5+10) = \frac{30}{5} = 6.$$

Pendiente y ordenada en el origen:

$$\hat{\beta}_2 = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}^2} = \frac{6}{2} = 3, \qquad \hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}} = 11 - 3 \cdot 3 = 2.$$

Recuperamos *exactamente* los coeficientes generadores $\beta_1 = 2$, $\beta_2 = 3$. Esto no es casualidad, pero tampoco es lo que ocurre en general: sucede porque, en este ejemplo pequeño, elegimos deliberadamente $\boldsymbol{u}$ de manera que su covarianza con $\boldsymbol{x}$ fuera cero (compruébese: $\frac{1}{5}\sum(x_i - \mu_{\boldsymbol{x}})u_i = \frac{1}{5}[(-2)(1) + (-1)(-2) + (0)(0) + (1)(2) + (2)(-1)] = \frac{1}{5}(-2+2+0+2-2) = 0$) y su media cero. Con datos reales, o incluso con datos simulados de la forma habitual (perturbaciones generadas al azar, sin imponer esta cancelación exacta), $\hat{\beta}_2$ y $\hat{\beta}_1$ se *aproximarán* a los valores generadores, pero no los reproducirán con exactitud —como puede comprobarse en la práctica de laboratorio, con su muestra de mayor tamaño y sus perturbaciones genuinamente aleatorias—.

Podemos además comprobar, con este mismo ejemplo, que los residuos coinciden con $\boldsymbol{u}$: como el ajuste recupera exactamente los coeficientes generadores, $\hat{e}_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_i) = y_i - (2+3x_i) = u_i$. Y, como debía ocurrir por construcción, $\sum \hat{e}_i = \sum u_i = 0$ y $\sum x_i \hat{e}_i = \sum x_i u_i = 0$ (la ortogonalidad con los regresores que impusimos al elegir $\boldsymbol{u}$).

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, ejemplo numérico análogo con datos de salario/educación.

8 Recapitulación y guiño a la sesión siguiente

Condición de ortogonalidad	Consecuencia
$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$	$\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}$;\; $\mu_{\widehat{\boldsymbol{y}}} = \mu_{\boldsymbol{y}}$
$\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$	$\hat{\beta}_2 = \sigma_{\boldsymbol{xy}}/\sigma_{\boldsymbol{x}}^2 = \rho_{\boldsymbol{xy}}\, \sigma_{\boldsymbol{y}}/\sigma_{\boldsymbol{x}}$

Nótese el caso límite $\hat{\beta}_2 = 0$: entonces $\hat{\beta}_1 = \mu_{\boldsymbol{y}}$, y recuperamos exactamente el ajuste de la lección 4 —la media—. La regresión simple generaliza aquella proyección sobre una recta añadiendo una segunda dirección; no la sustituye.

Dos ortogonalidades, dos coeficientes. Nada de matrices; solo Pitágoras y bilinealidad, aplicados dos veces.

Próxima sesión — Lección 8: las mismas dos ortogonalidades, combinadas con el teorema de Pitágoras, dan $\sigma_{\boldsymbol{y}}^2 = \sigma_{\widehat{\boldsymbol{y}}}^2 + \sigma_{\widehat{\boldsymbol{e}}}^2$. De ahí saldrá el $R^2 = \cos^2\theta$: ¿cómo de bueno es nuestro ajuste?

Recapitulación y guiño a la sesión siguiente

El recorrido de hoy ha sido, en el fondo, muy corto: partiendo de las dos condiciones de ortogonalidad que dejamos planteadas en la lección 6, y usando únicamente álgebra elemental (despejar, sustituir, dividir por n), hemos llegado a fórmulas cerradas para $\hat{\beta}_1$ y $\hat{\beta}_2$. Ninguna de las piezas era nueva: la media, la varianza, la covarianza y la correlación ya estaban completamente definidas y demostradas desde las lecciones 4 y 5. Lo único que hemos hecho hoy es plegar esas piezas sobre un problema de dos incógnitas.

Merece la pena señalar, aunque no lo desarrollemos aquí, que estas mismas fórmulas podrían obtenerse igualmente por la vía puramente analítica: minimizando $\|\widehat{\boldsymbol{e}}\|_s^2 = \frac{1}{n}\sum_i (y_i - \beta_1 - \beta_2 x_i)^2$ respecto a β_1, β_2 e igualando a cero las dos derivadas parciales. Esas dos condiciones de primer orden son, de hecho, exactamente las dos condiciones de ortogonalidad que hemos resuelto en esta lección —el mismo fenómeno que ya vimos, con un solo coeficiente, en la lección 4—:

1. La primera condición de ortogonalidad ($\widehat{\boldsymbol{e}} \perp \mathbf{1}$) fija el intercepto en función de la pendiente, $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}$, y garantiza que la recta ajustada pasa por el punto de las medias $(\mu_{\boldsymbol{x}}, \mu_{\boldsymbol{y}})$, así como que $\mu_{\widehat{\boldsymbol{y}}} = \mu_{\boldsymbol{y}}$.
2. La segunda condición ($\widehat{\boldsymbol{e}} \perp \boldsymbol{x}$) fija la pendiente como el cociente entre la covarianza y la varianza, $\hat{\beta}_2 = \sigma_{\boldsymbol{xy}}/\sigma_{\boldsymbol{x}}^2$, que puede reescribirse como la correlación reescalada por el cociente de las desviaciones típicas, $\hat{\beta}_2 = \rho_{\boldsymbol{xy}}\, \sigma_{\boldsymbol{y}}/\sigma_{\boldsymbol{x}}$.

También vimos un camino alternativo —cambiar a generadores ortogonales, $\mathbf{1}$ y $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$— que llega al mismo resultado sin resolver ningún sistema, proyectando cada coeficiente de forma independiente. Este segundo camino es el que se generalizará, con más de un regresor, en la lección 10 (regresión múltiple). Allí ya no bastarán dos condiciones de ortogonalidad, sino tantas como regresores tengamos; y la notación matricial —anunciada en la lección 5, pero todavía no utilizada— empezará a ganarse su lugar como forma compacta de escribir esto mismo.

La sesión de laboratorio que sigue a esta lección (regresión simple con datos reales) comprobará numéricamente, con un dataset completo, que efectivamente $\sum \hat{e}_i = 0$ y $\sum x_i \hat{e}_i = 0$ —las dos condiciones de ortogonalidad de hoy—, y que las fórmulas de $\hat{\beta}_1, \hat{\beta}_2$ coinciden con las que reporta Gretl automáticamente al pedir una regresión.

Y la próxima sesión teórica, la lección 8, dará el siguiente paso natural. Las mismas dos condiciones de ortogonalidad, vistas bajo la óptica del teorema de Pitágoras —con $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$ como suma de dos vectores ortogonales—, implican una descomposición de varianzas, $\sigma_{\boldsymbol{y}}^2 = \sigma_{\widehat{\boldsymbol{y}}}^2 + \sigma_{\widehat{\boldsymbol{e}}}^2$. De ella surgirá la medida de bondad de ajuste R^2, expresada geométricamente como el coseno al cuadrado de un ángulo. El ángulo de la lección 5 —el que definía la correlación— no ha desaparecido: solo cambia de protagonistas.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, secciones 2.2-2.3: resumen algebraico completo de $\hat{\beta}_1, \hat{\beta}_2$ y sus propiedades.
- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: proyección ortogonal sobre subespacios generados por vectores ortogonales entre sí.