Índice general

Lección 6. La regresión lineal simple como ajuste geométrico

Marcos Bujosa

19 de septiembre de 2026

Resumen

Generalizamos la proyección sobre una recta (la media, lección 4) a la proyección sobre un plano generado por dos vectores: $\mathbf{1}$ (de unos) y $\boldsymbol{x}$ (de datos). Todavía seguimos trabajando sin supuestos relacionados con la probabilidad: es el mismo problema geométrico de antes (buscar el vector más próximo a los datos $\boldsymbol{y}$), pero con una dirección de búsqueda adicional. Formalizamos aquí el término *regresor* (que ya asomó en la lección 3) y dejamos planteada, como tarea para la lección siguiente, la resolución del sistema que determina el ajuste.

"Después de una recta, un plano: la misma pregunta, un vector más."

- (slides) — (html) — (pdf)

1 De dónde venimos: Proyección ortogonal sobre una recta

- En la lección 4 proyectamos el vector de datos $\boldsymbol{y}$ sobre la **recta** $\mathcal{L}(\mathbf{1})$ buscando el múltiplo de $\mathbf{1}$ más próximo. El escalar por el que multiplicar $\mathbf{1}$ resultó ser la **media**.
- La pregunta era geométrica: ¿cuál es el múltiplo de $\mathbf{1}$ más cercano a $\boldsymbol{y}$?

En esta lección:

- Hacemos la misma pregunta, pero añadimos una *segunda dirección*.
- Proyectaremos sobre el **plano** generado por $\mathbf{1}$ y un segundo vector de datos $\boldsymbol{x}$.

Aún no hay variables aleatorias, ni perturbaciones, ni causalidad: sigue siendo pura geometría en $\mathbb{R}^n$.

De dónde venimos

En la lección 4 planteamos un problema muy concreto: dado un vector de datos $\boldsymbol{y} \in \mathbb{R}^n$, ¿cuál es el múltiplo de $\mathbf{1}$ (el vector constante) más próximo a $\boldsymbol{y}$? La respuesta es $\mu_{\boldsymbol{y}}\mathbf{1}$ (la media de los datos multiplicada por $\mathbf{1}$). Este resultado es la **proyección ortogonal** de $\boldsymbol{y}$ sobre $\mathcal{L}(\mathbf{1})$, es decir, sobre la recta en $\mathbb{R}^n$ que pasa por el origen y contiene a todos los múltiplos de $\mathbf{1}$.

En esta lección hacemos la misma pregunta, pero con una importante diferencia: en lugar de disponer de un único vector generador ($\mathbf{1}$), disponemos de *dos*. Además de $\mathbf{1}$, tenemos ahora un

segundo vector de datos $\boldsymbol{x} \in \mathbb{R}^n$ (por ejemplo, la superficie de n viviendas, si $\boldsymbol{y}$ es el precio). La pregunta pasa a ser: ¿cuál es la combinación de $\mathbf{1}$ y $\boldsymbol{x}$ más próxima a $\boldsymbol{y}$?

Quiero subrayar algo importante desde el principio: *no vamos a hablar todavía de modelos, de supuestos, ni de si $\boldsymbol{x}$ "causa" $\boldsymbol{y}$*. Eso vendrá más adelante, cuando dispongamos de herramientas para plantearlo con rigor. Por ahora, igual que con la media, esto es un problema puramente geométrico: tenemos unos datos $\boldsymbol{x}$ y queremos encontrar la combinación de $\mathbf{1}$ y $\boldsymbol{x}$ (un punto en un cierto subespacio) que está más cerca de $\boldsymbol{y}$.

Para profundizar:

- Bujosa, M. (2024). *Curso de Álgebra Lineal*, cap. 11. enlace.

2 El plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$

- Un vector generador $\boldsymbol{a} \Rightarrow \mathcal{L}(\boldsymbol{a})$: una *recta*.
- Dos vectores generadores $\boldsymbol{a}, \boldsymbol{b}$ (no alineados) $\Rightarrow \mathcal{L}(\boldsymbol{a}, \boldsymbol{b})$: un *plano*.

$$\mathcal{L}(\mathbf{1}, \boldsymbol{x}) = \{\beta_1 \mathbf{1} + \beta_2 \boldsymbol{x} \mid \beta_1, \beta_2 \in \mathbb{R}\}$$

El plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$

Recordemos de la lección 2 que una *combinación lineal* de dos vectores $\boldsymbol{a}$ y $\boldsymbol{b}$ es cualquier expresión de la forma $\beta_1 \boldsymbol{a} + \beta_2 \boldsymbol{b}$, con β_1, β_2 números reales cualesquiera. El conjunto de *todas* las combinaciones lineales posibles de $\boldsymbol{a}$ y $\boldsymbol{b}$ se denota $\mathcal{L}(\boldsymbol{a}, \boldsymbol{b})$, generalizando la notación $\mathcal{L}(\boldsymbol{a})$ ya introducida en la lección 4 para el caso de un único vector generador.

Cuando $\boldsymbol{a}$ y $\boldsymbol{b}$ no son múltiplos el uno del otro (es decir, cuando no están alineados dentro de la misma recta), el conjunto $\mathcal{L}(\boldsymbol{a}, \boldsymbol{b})$ es, geométricamente, un *plano* que pasa por el origen: contiene, de manera informal, *dos direcciones* independientes (retomamos aquí, sin formalizarla, la noción de dimensión que ya usamos en las lecciones 4 y 5). Cualquier punto del plano se alcanza combinando adecuadamente los "desplazamientos" o "pasos" necesarios en esas dos direcciones.

En nuestro caso, los dos vectores generadores son $\mathbf{1}$ (el vector de unos) y $\boldsymbol{x}$ (nuestro segundo vector de datos). El plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ contiene, en particular, la recta $\mathcal{L}(\mathbf{1})$ con la que ya trabajamos en la lección 4 (es el caso particular $\beta_2 = 0$), pero es un conjunto mucho más rico: contiene todas las variaciones producidas al sumarle a cualquier múltiplo de $\mathbf{1}$ otro múltiplo de $\boldsymbol{x}$.

Nótese que si $\boldsymbol{x}$ fuera un múltiplo de $\mathbf{1}$ (es decir, si todos los datos de $\boldsymbol{x}$ fueran idénticos o, dicho de otro modo, si la varianza de $\boldsymbol{x}$ fuera cero), $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ colapsaría de nuevo en una simple recta: no ganaríamos ninguna dirección nueva. Volveremos sobre este caso degenerado en la lección 14, dedicada a la colinealidad; por ahora basta con tenerlo presente como advertencia.

Para profundizar:

- Strang, G. (2016). *Introduction to Linear Algebra* (5ª ed.), sección 1.1: combinaciones lineales de dos vectores y el plano que generan.
- Bujosa, M. *Curso de Álgebra Lineal*, capítulo 11: libro online.

- Vídeo: Linear combinations, span, and basis vectors de 3Blue1Brown (capítulo 2 de *Essence of Linear Algebra*). Doce minutos sobre qué conjunto generan uno, dos o tres vectores.

3 Un matiz importante: $\mathbf{1}$ y $\boldsymbol{x}$ no son ortogonales

Recordamos de la lección 4:

$$\mu_{\boldsymbol{x}} = \langle \boldsymbol{x}, \mathbf{1} \rangle_s.$$

Si $\mu_{\boldsymbol{x}} \neq 0$ (el caso habitual), entonces $\mathbf{1}$ y $\boldsymbol{x}$ *no son ortogonales*.

Los ejes $\mathcal{L}(\mathbf{1})$ y $\mathcal{L}(\boldsymbol{x})$ generalmente se cortan *oblicuamente*, no en ángulo recto.

Un matiz importante: $\mathbf{1}$ y $\boldsymbol{x}$ no son ortogonales

Este detalle técnico va a condicionar la forma en que dibujemos —y, sobre todo, la forma en que *interpretemos* — la figura de la siguiente transparencia.

En la lección 4 vimos que $\mu_{\boldsymbol{x}} = \langle \boldsymbol{x}, \mathbf{1} \rangle_s$ es el producto escalar entre $\boldsymbol{x}$ y $\mathbf{1}$. Y en la lección 3 aprendimos que dos vectores son ortogonales exactamente cuando su producto escalar es cero. Por tanto, si la media de $\boldsymbol{x}$ no es cero —que es una situación habitual con datos económicos: precios, superficies, años de educación, todos ellos con media positiva— entonces $\boldsymbol{x}$ y $\mathbf{1}$ *no* son ortogonales.

¿Qué consecuencia tiene esto? Que si dibujamos el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ usando como ejes de referencia $\mathbf{1}$ y $\boldsymbol{x}$ directamente, esos dos ejes no se cortarán en ángulo recto, sino oblicuamente. Esto es exactamente lo que verá en la figura de la próxima transparencia.

Fíjese, no obstante, que podemos generar *el mismo plano* con otro par de vectores que sí sean ortogonales entre sí. En concreto,

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

donde $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$ es lo que en la lección 4 llamamos el *vector en desviaciones* de $\boldsymbol{x}$. Este vector en desviaciones SÍ es ortogonal a $\mathbf{1}$ (es, precisamente, la proyección de $\boldsymbol{x}$ sobre $\mathcal{L}(\mathbf{1})^{\perp}$, el conjunto de todos los vectores ortogonales a $\mathbf{1}$). El plano no cambia —es el mismo conjunto de puntos—, pero si lo dibujamos usando este nuevo par de generadores, los ejes sí aparecerían perpendiculares.

Por ahora nos quedamos con la versión "cruda": trabajaremos con $\mathbf{1}$ y $\boldsymbol{x}$ tal cual, sin centrar. Esto tiene la ventaja de que el ajuste resultante $(\hat{\beta}_1, \hat{\beta}_2)$ es directamente el que buscamos, sin pasos intermedios. Pero conviene tener en mente esta alternativa "centrada",[1] porque reaparecerá cuando, en lecciones posteriores, discutamos cómo se relaciona la pendiente de la regresión con la correlación entre $\boldsymbol{x}$ e $\boldsymbol{y}$.

Para profundizar:

- Wooldridge, J. M. (2020). *Introductory Econometrics*, Apéndice C (repaso del efecto de transformaciones lineales).

[1] recuerde que en estadística "/centrar/" un vector de datos significa restar la media a cada dato.

4 El problema de ajuste

Buscamos $\hat{\beta}_1, \hat{\beta}_2$ tales que

$$\|\boldsymbol{y} - (\beta_1 \mathbf{1} + \beta_2 \boldsymbol{x})\|_s$$

sea *mínima*.

$$\widehat{\boldsymbol{y}} = \hat{\beta}_1 \mathbf{1} + \hat{\beta}_2 \boldsymbol{x} \quad \text{(el ajuste: la proyección)}$$

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \quad \text{(el residuo o } \textit{vector de residuos} \text{ del ajuste)}$$

El problema de ajuste

Formalicemos el problema que veníamos planteando de manera informal. Dado el vector de datos $\boldsymbol{y}$ y el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, queremos encontrar los dos números $\hat{\beta}_1$ y $\hat{\beta}_2$ (nótese: son *escalares*, sin negrita, exactamente como el α que buscábamos en la lección 4 cuando solo había un vector generador) que hacen que la distancia (norma estadística) entre $\boldsymbol{y}$ y el punto $\beta_1 \mathbf{1} + \beta_2 \boldsymbol{x}$ del plano sea la menor posible.

Al vector del plano que logra esa distancia mínima lo llamamos el *ajuste* MCO[2] y lo denotamos $\widehat{\boldsymbol{y}}$. Es importante no confundir el vector $\widehat{\boldsymbol{y}}$ con los escalares $\hat{\beta}_1, \hat{\beta}_2$ que lo determinan: quien *es* la proyección ortogonal es el vector $\widehat{\boldsymbol{y}}$; los coeficientes $\hat{\beta}_1, \hat{\beta}_2$ son, simplemente, los números que hay que usar para construirlo a partir de $\mathbf{1}$ y $\boldsymbol{x}$.

A la diferencia entre los datos observados y el ajuste, $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$, la llamamos *residuo* o vector con los *errores de ajuste*. Aquí introducimos una precisión terminológica: reservamos la palabra "residuo" exclusivamente para este contexto, el de un ajuste con regresores no constantes. Cuando en la lección 4 hablábamos de $\boldsymbol{y} - \mu_{\boldsymbol{y}} \mathbf{1}$, usábamos deliberadamente la expresión "vector en desviaciones", no "residuo" — comparten una construcción geométrica parecida (una diferencia entre el dato y un ajuste), pero mantenemos los nombres distintos porque corresponden a ajustes distintos, y porque en las próximas lecciones el residuo de la regresión jugará un papel muy específico (será la pieza clave para estimar la variabilidad no explicada al ajustar un modelo).

Para profundizar:

- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 2, sección 2.2: la obtención de las estimaciones MCO, planteada allí como minimización de una suma de cuadrados (la misma que aquí planteamos como distancia mínima).
- Strang, G. (2016), sección 4.3: *Least squares approximations*, el mismo problema desde el álgebra lineal.

5 La geometría del ajuste

Versión interactiva: cuaderno 2 en MyBinder, parte 1 (el plano, el ajuste y el residuo, rotables).

[2] El ajuste de Mínimos Cuadrados Ordinarios; en inglés OLS (Ordinary Least Squares)

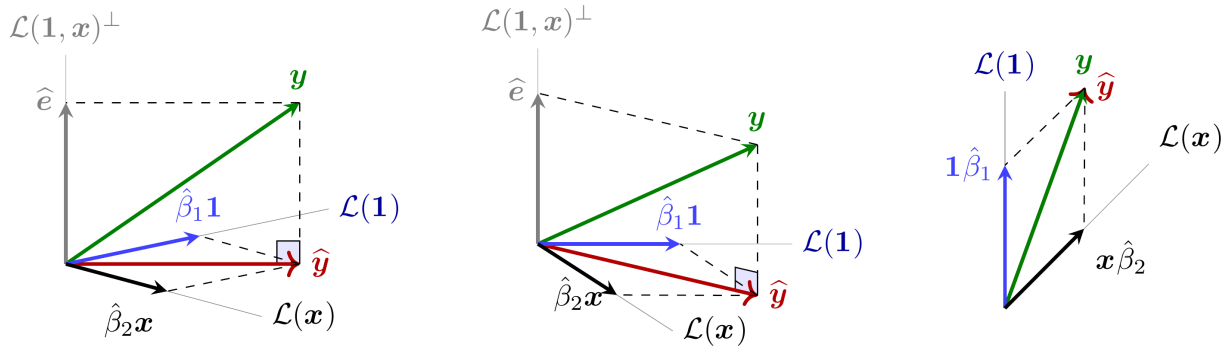

Figura 1: Proyección ortogonal de $\boldsymbol{y}$ sobre el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, vista desde tres ángulos: (izquierda) de frente al plano $(\widehat{\boldsymbol{y}}, \widehat{\boldsymbol{e}})$; (centro) de frente al plano $(\mathcal{L}(\mathbf{1}), \mathcal{L}(\mathbf{1}, \boldsymbol{x})^\perp)$; (derecha) vista cenital, perpendicular al plano de los regresores.

La geometría del ajuste

Detengámonos con calma en esta figura, porque condensa toda la geometría de la lección.

La escena representa un caso con $n = 3$ datos (por eso podemos dibujarla en un espacio tridimensional) pero la podemos leer como una representación esquemática (no literal) del caso general en $\mathbb{R}^n$ (para cualquier cantidad de datos $n \geq 3$). Tenemos el vector de datos $\boldsymbol{y}$ (en verde), y el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, generado por los dos vectores $\mathbf{1}$ (contenido en el eje $\mathcal{L}(\mathbf{1})$) y $\boldsymbol{x}$ (contenido en el eje $\mathcal{L}(\boldsymbol{x})$).[3] Como $\boldsymbol{x}$ tiene media distinta de cero, estos dos ejes no son perpendiculares entre sí, sino oblicuos — exactamente lo que anticipamos en la transparencia anterior.

El vector rojo, $\widehat{\boldsymbol{y}}$, es la proyección ortogonal de $\boldsymbol{y}$ sobre ese plano: es el punto del plano más próximo a $\boldsymbol{y}$. De la punta de $\widehat{\boldsymbol{y}}$ salen líneas discontinuas, paralelas a los ejes, que llegan hasta sus componentes sobre cada uno de los dos generadores: $\hat{\beta}_1\mathbf{1}$ (en azul) y $\hat{\beta}_2\boldsymbol{x}$ (en negro). Esta es la forma geométrica de leer que $\widehat{\boldsymbol{y}} = \hat{\beta}_1\mathbf{1} + \hat{\beta}_2\boldsymbol{x}$: sumando esas dos componentes se reconstruye $\widehat{\boldsymbol{y}}$.

De la punta de $\widehat{\boldsymbol{y}}$ sale también una línea discontinua vertical que llega hasta $\boldsymbol{y}$, y en la esquina de esa línea con el plano aparece marcado un ángulo recto para recordarnos que estamos ante la proyección ortogonal de $\boldsymbol{y}$ sobre el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$. En el eje vertical aparece el vector $\widehat{\boldsymbol{e}}$ (en gris), que resulta ser ortogonal a todo el plano. El eje vertical sobre el que se encuentra está marcado con $\mathcal{L}(\mathbf{1}, \boldsymbol{x})^\perp$ (es el conjunto de todos los vectores ortogonales al plano de los regresores).

Las tres vistas de la figura nos ayudan a ver distintos aspectos del ajuste:

- La vista de la *izquierda* está tomada de frente al plano formado por $\widehat{\boldsymbol{y}}$ y $\widehat{\boldsymbol{e}}$, cuya suma es el vector de datos $\boldsymbol{y}$. Desde aquí se aprecia con claridad un rectángulo dividido en dos triángulos rectángulos cuya hipotenusa es $\boldsymbol{y}$: la misma estructura pitagórica que ya usamos en la lección 5, pero ahora con $\widehat{\boldsymbol{y}}$ en el papel que allí jugaba el vector de medias.

[3] Los vectores $\mathbf{1}$ y $\boldsymbol{x}$ no están explícitamente representados en la figura para no complicarla más.

- La vista *central* está tomada de frente al plano formado por el eje $\mathcal{L}(\mathbf{1})$ y el eje vertical $\mathcal{L}(\mathbf{1}, \boldsymbol{x})^{\perp}$ y permite apreciar que los regresores no son perpendiculares entre sí.
- La vista de la *derecha* es una vista *cenital*: mirando perpendicularmente al plano de los regresores, desde arriba. Es aquí donde mejor se aprecia la oblicuidad entre los ejes $\mathcal{L}(\mathbf{1})$ y $\mathcal{L}(\boldsymbol{x})$: si proyecta mentalmente el vector $\hat{\beta}_2\boldsymbol{x}$ sobre el eje $\mathcal{L}(\mathbf{1})$ (que en esta vista se ve vertical) y suma esa proyección a $\hat{\beta}_1\mathbf{1}$, llega al mismo punto donde caen las proyecciones tanto de $\widehat{\boldsymbol{y}}$ como de $\boldsymbol{y}$ sobre esa recta. Esto es una consecuencia directa de que $\widehat{\boldsymbol{e}}$ es ortogonal a $\mathcal{L}(\mathbf{1})$, y anticipa una propiedad que veremos en la lección 7: la media de $\widehat{\boldsymbol{y}}$ coincide con la media de $\boldsymbol{y}$.

Nótese lo que esta figura NO contiene: ningún supuesto sobre cómo se generaron los datos, ninguna variable aleatoria, ninguna perturbación "poblacional". Es geometría pura sobre un vector de datos y un plano.

Para profundizar:

- Strang, G. (2016), sección 4.2: *Projections*, con la proyección sobre un subespacio de dimensión mayor que uno.
- Bujosa, M. *Curso de Álgebra Lineal*, capítulo 11: libro online.

6 Lo que la ortogonalidad nos va a dar

$\widehat{\boldsymbol{e}}$ es ortogonal a *todo* el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$.

En particular, es ortogonal a cada uno de sus generadores:

$$\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0 \qquad \text{(el residuo } \widehat{\boldsymbol{e}} \text{ tiene media cero)}$$

y

$$\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0 \quad (\text{junto con } \langle \widehat{e}, \mathbf{1} \rangle_s = 0) \implies \widehat{\boldsymbol{e}} \text{ tiene covarianza cero con } \boldsymbol{x}.$$

Dos ecuaciones, dos incógnitas $(\hat{\beta}_1, \hat{\beta}_2)$... *(la próxima lección resuelve el sistema).*

Lo que la ortogonalidad nos va a dar

Esta transparencia es, en cierto sentido, el destino de toda la lección, aunque dejemos su desarrollo completo para la siguiente.

La propiedad definitoria de una proyección ortogonal —ya la usamos en la lección 4 con un único vector generador— es que el residuo es ortogonal a *todo* el subespacio sobre el que proyectamos, no solo a alguno de sus vectores. En la lección 4, como el subespacio era una simple recta $\mathcal{L}(\mathbf{1})$, esto se traducía en una única condición: $\langle (\boldsymbol{y} - \mu_{\boldsymbol{y}}\mathbf{1}), \mathbf{1} \rangle_s = 0$, de la cual salía la fórmula para calcular la media $\mu_{\boldsymbol{y}}$.

Ahora que el subespacio es un plano generado por *dos* vectores, la misma exigencia —que $\widehat{\boldsymbol{e}}$ sea ortogonal a todo el plano— se traduce, de manera natural, en *dos* condiciones: $\widehat{\boldsymbol{e}}$ debe ser ortogonal tanto a $\mathbf{1}$ como a $\boldsymbol{x}$ (los dos generadores del plano): basta con que un vector sea ortogonal a los

generadores de un subespacio para que sea ortogonal a *cualquier* combinación lineal de ellos (y por tanto, a todo el subespacio).

Merece la pena leer con detenimiento qué significa cada condición por separado:

- $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ es exactamente la misma condición que ya vimos en la lección 4 para la media: dice que la media de las componentes de $\widehat{\boldsymbol{e}}$ es cero, es decir, $\sum_i \hat{e}_i = 0$. Por tanto, el ajuste $\widehat{\boldsymbol{y}}$, sea cual sea, no puede sistemáticamente sobrestimar ni subestimar los datos $\boldsymbol{y}$ en promedio (pues la suma de los errores necesariamente se cancela).

- $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$ es una condición *nueva*, que no tenía sentido cuando solo disponíamos de $\mathbf{1}$.

  En conjunción con la primera condición (que $\widehat{\boldsymbol{e}}$ es ortogonal a los vectores constantes, y por tanto ya está en desviaciones), arroja una propiedad algebraica sobre la que merece la pena detenerse, pues la usaremos repetidas veces. Si calculamos el producto escalar de $\widehat{\boldsymbol{e}}$ con otro vector en desviaciones respecto a su media, aplicando las propiedades tenemos

$$\sigma_{\widehat{\boldsymbol{e}}\boldsymbol{x}} \quad = \quad \Big\langle \widehat{\boldsymbol{e}}, (\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}) \Big\rangle_s \quad = \quad \langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s - \underbrace{\Big\langle \widehat{\boldsymbol{e}}, (\mu_{\boldsymbol{x}}\mathbf{1}) \Big\rangle_s}_{=0;\ \widehat{\boldsymbol{e}} \perp \mathcal{L}(\mathbf{1})} \quad = \quad \langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s.$$

  ¡Este es un principio general que será muy útil! *Si un vector tiene media cero, su producto escalar con cualquier otro vector resulta ser la covarianza.*

  Retomando el vocabulario de la lección 5: cuando $\widehat{\boldsymbol{e}}$ ya está centrado (cuando su media es nula), decir que $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0 = \frac{1}{n} \sum \widehat{e}_i \, x_i$ es lo mismo que decir que $\widehat{\boldsymbol{e}}$ y $\boldsymbol{x}$ tienen **covarianza cero**. Por tanto, *cuando el residuo no es nulo*[4], esto equivale a que $\widehat{\boldsymbol{e}}$ y $\boldsymbol{x}$ están **incorrelados** (correlación cero).

Recapitulando, tenemos dos ecuaciones y dos incógnitas ($\hat{\beta}_1$ y $\hat{\beta}_2$). En principio esto debería bastar para determinar $\hat{\beta}_1$ y $\hat{\beta}_2$ de forma única[5] —tal como ocurría con una sola ecuación y una sola incógnita en la lección 4—. Resolver *explícitamente* ese sistema, y obtener fórmulas cerradas para $\hat{\beta}_1$ y $\hat{\beta}_2$ en función de los datos, es precisamente la tarea de la lección 7.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, sección 2.2: las condiciones de primer orden del problema de mínimos cuadrados, que son estas mismas dos ortogonalidades escritas como sumatorios.

7 Vocabulario: regresor

Definición (regresor). Llamamos *regresor* a cada uno de los vectores que entran en la combinación lineal del ajuste (cada vector que multiplica a un coeficiente $\hat{\beta}_j$).

En la regresión lineal simple hay *dos* regresores: $\mathbf{1}$ y $\boldsymbol{x}$.

[4] y el vector en desviaciones $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$ tampoco es nulo; recuerde que para que la correlación esté definida, las desviaciones típicas (las normas) que dividen a la covarianza no pueden ser cero.

[5] Siempre y cuando $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ sea un plano.

(Postponemos deliberadamente el término "variable explicativa": ese término trae consigo un matiz causal que aún no estamos en condiciones de justificar.)

Vocabulario: regresor

Formalizamos aquí, de forma deliberadamente aséptica, el término que va a acompañarnos durante el resto del curso: *regresor*. Un regresor es, simplemente, cualquiera de los vectores que aparecen multiplicados por un coeficiente $\hat{\beta}_j$ en la combinación lineal que estamos ajustando. En el modelo de esta lección hay exactamente dos: el vector de unos $\mathbf{1}$ y el vector de datos $\boldsymbol{x}$.

Es importante no confundir "regresor" con "variable explicativa". Un regresor es: un vector concreto que entra en la combinación lineal. Una variable explicativa, en cambio, es un *concepto sustantivo* (por ejemplo, "la educación de una persona", o "la superficie de una vivienda"), que puede entrar en el modelo mediante uno o varios regresores. Por ejemplo, si quisiéramos explicar el salario usando tanto los años de educación como el cuadrado de dichos años, tendríamos *dos* regresores (la columna de años de educación y la columna con esos años al cuadrado), pero seguiría siendo *una sola* variable explicativa (la educación), que entra en el modelo a través de dos columnas distintas.

Por ahora, con solo dos regresores ($\mathbf{1}$ y $\boldsymbol{x}$), esta distinción es casi anecdótica. Pero conviene fijar el vocabulario desde ahora, porque en lecciones posteriores construiremos modelos con varios regresores a la vez, y esta distinción entre "columna de la matriz de regresores" y "concepto económico involucrado" será relevante para leer correctamente los resultados.

Nótese, además, que en esta lección "regresor" es un término puramente *estructural*, sin ninguna connotación causal: decir que $\boldsymbol{x}$ es un regresor no presupone que $\boldsymbol{x}$ "cause" $\boldsymbol{y}$, ni ninguna otra relación de tipo teórico. Es, sencillamente, uno de los vectores generadores del plano sobre el que proyectamos. La discusión sobre qué podemos —y qué NO podemos— concluir causalmente a partir de un ajuste como este llegará más adelante, cuando dispongamos de más herramientas con las que plantear esa pregunta con más rigor.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, sección 2.1: la terminología habitual (variable dependiente e independiente, regresando y regresor) y sus sinónimos en la literatura.

8 Recapitulación y guiño a la lección siguiente

Qué hemos hecho:

- Generalizado "proyección sobre una recta" (lección 4) a "proyección sobre un plano".
  - La media: el caso con *un solo* regresor ($\mathbf{1}$); ahora generalizamos a *dos*.
- Visto que $\mathbf{1}$ y $\boldsymbol{x}$ no son ortogonales en general (es decir, normalmente $\mu_{\boldsymbol{x}} \neq 0$).
- Identificado *dos* condiciones de ortogonalidad: $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ y $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$.
- Fijado el vocabulario: *regresor*.

Qué falta (Lección 7):

- Resolver el sistema: fórmulas cerradas para $\hat{\beta}_1$ y $\hat{\beta}_2$.

"Dos ecuaciones de ortogonalidad, dos incógnitas: solo falta determinarlas."

Recapitulación y guiño a la lección siguiente

Recapitulemos el camino recorrido en esta lección. Partimos de una pregunta de la lección 4 —¿cuál es la combinación más próxima a nuestros datos $\boldsymbol{y}$?— y la generalizamos añadiendo un segundo vector generador, $\boldsymbol{x}$. Vimos que esto nos lleva de una recta, $\mathcal{L}(\mathbf{1})$, a un plano, $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$.

Vimos, con ayuda de la figura de tres vistas, cómo se organiza geométricamente el ajuste: el vector $\boldsymbol{y}$ (verde) se descompone en dos componentes ortogonales: $\widehat{\boldsymbol{y}}$ (rojo) en el plano y el vector con los errores de ajuste $\widehat{\boldsymbol{e}}$ (gris) que es perpendicular a dicho plano. Además, el vector $\widehat{\boldsymbol{y}}$ (la proyección, roja en la figura) se descompone en sus dos componentes sobre los ejes $\mathcal{L}(\mathbf{1})$ y $\mathcal{L}(\boldsymbol{x})$ del plano.

Y llegamos al resultado central, que dejamos planteado pero sin resolver: esa perpendicularidad de los residuos $\widehat{\boldsymbol{e}}$ respecto del plano se traduce en *dos* condiciones de ortogonalidad simultáneas, $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ y $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$ — dos ecuaciones para dos incógnitas, $\hat{\beta}_1$ y $\hat{\beta}_2$.

En la lección 7 resolveremos explícitamente ese sistema. Veremos que la fórmula que se obtiene para $\hat{\beta}_2$ (la pendiente) tiene una lectura muy natural en términos de lo que ya sabemos de la lección 5: guarda una relación directa con la covarianza entre $\boldsymbol{x}$ e $\boldsymbol{y}$, y con la correlación $\rho_{\boldsymbol{xy}}$. Y la fórmula para $\hat{\beta}_1$ (la constante) se apoyará, a su vez, en las medias de $\boldsymbol{x}$ e $\boldsymbol{y}$ — de manera que, en cierto sentido, todo lo que construimos en las lecciones 4 y 5 sobre medias, varianzas, covarianzas y correlaciones reaparecerá aquí, ahora como ingredientes de un ajuste más rico.

Como siempre, quiero insistir en el punto de partida de esta lección: todo lo que hemos hecho es geometría pura sobre un vector de datos concreto, $\boldsymbol{y}$, y dos vectores generadores concretos, $\mathbf{1}$ y $\boldsymbol{x}$. No hemos necesitado hablar de modelos poblacionales, ni de supuestos sobre cómo se generaron los datos, ni de si $\boldsymbol{x}$ "causa" $\boldsymbol{y}$. Esa discusión —necesaria para poder *interpretar* el ajuste— llegará más adelante en el curso, cuando dispongamos de las herramientas adecuadas para plantearla con cuidado.

Para profundizar:

- Strang, G. (2016). *Introduction to Linear Algebra*, cap. 4 (mínimos cuadrados como proyección).