Índice general

Lección 5. Varianza, covarianza y correlación

Marcos Bujosa

19 de septiembre de 2026

Resumen

La varianza como el cuadrado de una norma, la covarianza como un producto escalar y la correlación como el coseno de un ángulo. El teorema de Pitágoras de la lección 4, ahora con dos vectores.

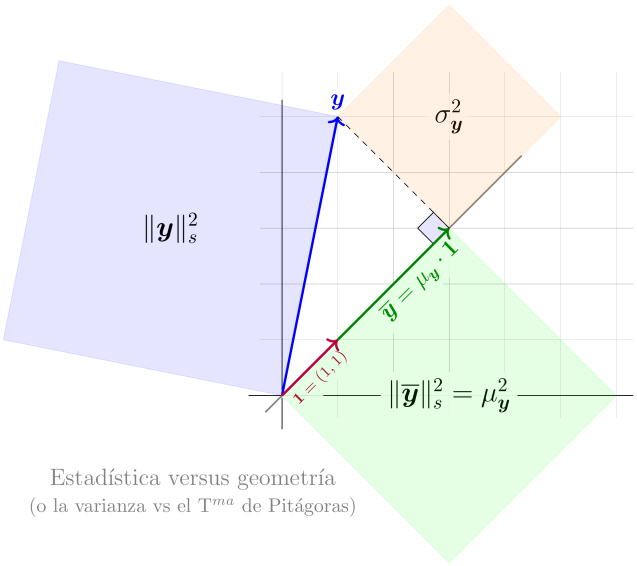
"Dos vectores, un ángulo entre sus componentes no constantes: ahí vive toda la correlación."

- (slides) — (html) — (pdf)

1 De dónde venimos: Pitágoras, otra vez

- $\mu_{\boldsymbol{y}}$: escalar que multiplica a $\mathbf{1}$ para obtener la proyección $\overline{\boldsymbol{y}}$ sobre $\mathcal{L}(\mathbf{1})$ (recordemos: $\mu_{\boldsymbol{y}} = \langle \boldsymbol{y}, \mathbf{1} \rangle_s$, la media ES ese producto escalar).
- $\sigma_{\boldsymbol{y}}$: norma del vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$ (perpendicular a $\mathbf{1}$).

En la lección 4 apareció la identidad: $\|\boldsymbol{y}\|_s^2 = \|\overline{\boldsymbol{y}}\|_s^2 + \|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 = \|\overline{\boldsymbol{y}}\|_s^2 + \sigma_{\boldsymbol{y}}^2$.



Estadística versus geometría
(o la varianza vs el T^{ma} de Pitágoras)

Tres preguntas:

- ¿Podemos calcular la varianza sin haber calculado $\boldsymbol{y} - \overline{\boldsymbol{y}}$?

Y si consideramos *dos* vectores de datos, $\boldsymbol{x}$ e $\boldsymbol{y}$:

- ¿Qué mide el producto escalar de sus *vectores en desviaciones*? (la covarianza)
- ¿Qué mide el coseno del ángulo entre esos dos *vectores en desviaciones*? (la correlación)

Mismo escenario que en la lección 4: proyecciones, normas y ángulos en $\mathcal{L}(\mathbf{1})^{\perp}$.

De dónde venimos: Pitágoras, otra vez

La lección 4 terminó con una identidad que conviene tener muy presente: el vector de datos $\boldsymbol{y}$ se descompone en *dos componentes ortogonales*: su proyección sobre $\mathcal{L}(\mathbf{1})$ (el vector de medias $\overline{\boldsymbol{y}} = \mu_{\boldsymbol{y}}\mathbf{1}$) que es la *componente constante*, y su vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$ (perpendicular a $\mathbf{1}$) que es la *componente variable*. El teorema de Pitágoras, aplicado a esa descomposición, dio

$$\|\boldsymbol{y}\|_s^2 = \|\overline{\boldsymbol{y}}\|_s^2 + \sigma_{\boldsymbol{y}}^2.$$

Hoy vamos a explorar esa misma geometría un paso más allá y en dos direcciones.

Primero, vamos a mirar con más detalle la propia varianza $\sigma_{\boldsymbol{y}}^2$ (el cuadrado de la norma de la componente variable), porque de ahí sale una fórmula de cálculo muy conocida en estadística descriptiva.

Segundo —y esto es lo verdaderamente nuevo—, vamos a dejar de trabajar con un único vector y vamos a contemplar *dos* vectores de datos simultáneamente, $\boldsymbol{x}$ e $\boldsymbol{y}$. Cuando tenemos dos vectores, dos preguntas geométricas naturales son: ¿cuánto vale el producto escalar de sus vectores en desviaciones (sus componentes variables)? y ¿qué ángulo forman dichas componentes variables? Las respuestas son, respectivamente, la *covarianza* y la *correlación*.

Para profundizar:

- Wooldridge, J. *Introductory Econometrics* (2020), Apéndice C: repaso de varianza, covarianza y correlación desde la estadística clásica (sin la lectura geométrica que estamos construyendo aquí).

2 La varianza: el cuadrado de la norma

Definición. La *varianza* de $\boldsymbol{y}$ es el cuadrado de la norma del vector en desviaciones respecto a la media $\boldsymbol{y} - \overline{\boldsymbol{y}}$:

$$\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 \;=\; \sigma_{\boldsymbol{y}}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_{\boldsymbol{y}})^2.$$

No es un concepto verdaderamente nuevo: es, literalmente, elevar al cuadrado la norma $\sigma_{\boldsymbol{y}}$ que ya conocíamos de la lección 4.

De la identidad de Pitágoras, despejando $\sigma_{\boldsymbol{y}}^2$ tenemos:

$$\|\boldsymbol{y}\|_s^2 = \|\overline{\boldsymbol{y}}\|_s^2 + \sigma_{\boldsymbol{y}}^2 \implies \sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y}\|_s^2 - \|\overline{\boldsymbol{y}}\|_s^2.$$

¿Qué son exactamente $\|\boldsymbol{y}\|_s^2$ y $\|\overline{\boldsymbol{y}}\|_s^2$ en términos de sumatorios o, mejor aún, en términos de medias aritméticas?

La varianza: el cuadrado de la norma

En la lección 4 llamamos $\sigma_{\boldsymbol{y}}$ a la *norma* del vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$. La *varianza* no es un objeto distinto: es simplemente el cuadrado de esa norma,

$$\sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 = \langle \boldsymbol{y} - \overline{\boldsymbol{y}},\, \boldsymbol{y} - \overline{\boldsymbol{y}} \rangle_s = \frac{1}{n}\sum_{i=1}^{n} (y_i - \mu_{\boldsymbol{y}})^2.$$

Esta es la definición habitual de la varianza en estadística descriptiva. Aquí insistimos en que no aparece de la nada: es el cuadrado de una norma, y esa norma es la que mide la distancia de $\boldsymbol{y}$ al vector constante más próximo, $\overline{\boldsymbol{y}}$; o, visto de otra manera, es la longitud de la componente variable de $\boldsymbol{y}$.

La identidad de Pitágoras de la lección 4, $\|\boldsymbol{y}\|_s^2 = \|\overline{\boldsymbol{y}}\|_s^2 + \sigma_{\boldsymbol{y}}^2$, puede despejarse para obtener

$$\sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y}\|_s^2 - \|\overline{\boldsymbol{y}}\|_s^2.$$

Esta forma alternativa de escribir la varianza —la que en los cursos de estadística descriptiva se suele presentar como "fórmula de cálculo" o "fórmula abreviada"— no es un truco algebraico: es lo que se obtiene del teorema de Pitágoras si se despeja la varianza. Para poder calcularla solo falta saber qué es $\|\boldsymbol{y}\|_s^2$ en forma de sumatorios o de medias aritméticas; y eso es justamente lo que hace la siguiente transparencia.

3 Varianza y Pitágoras

Los tres lados del triángulo rectángulo, elevados al cuadrado:

$$\underbrace{\|\boldsymbol{y}\|_s^2}_{\text{hipotenusa}^2} = \underbrace{\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s^2}_{\text{cateto}^2} + \underbrace{\sigma_{\boldsymbol{y}}^2}_{\text{cateto}^2}.$$

Como $\|\mathbf{1}\|_s = 1$: $\quad \|\mu_{\boldsymbol{y}}\mathbf{1}\|_s = |\mu_{\boldsymbol{y}}| \cdot \|\mathbf{1}\|_s = |\mu_{\boldsymbol{y}}| \implies \|\mu_{\boldsymbol{y}}\mathbf{1}\|_s^2 = \mu_{\boldsymbol{y}}^2$.

Despejando $\sigma_{\boldsymbol{y}}^2$ y recordando que $\|\boldsymbol{y}\|_s^2 = \langle \boldsymbol{y}, \boldsymbol{y} \rangle_s$:

$$\boxed{\sigma_{\boldsymbol{y}}^2 \;=\; \|\boldsymbol{y}\|_s^2 - \mu_{\boldsymbol{y}}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i^2 \;-\; \mu_{\boldsymbol{y}}^2 \;=\; \mu_{\boldsymbol{y}^2} \;-\; \mu_{\boldsymbol{y}}^2;}$$

donde $\boldsymbol{y}^2$ denota el vector cuyas componentes son el cuadrado de las de $\boldsymbol{y}$.

La "fórmula de cálculo" de la varianza de los cursos de estadística no es truco algebraico: es una implicación del Tma. de Pitágoras.

Varianza y Pitágoras

Si nos fijamos en la figura de la transparencia "De dónde venimos: Pitágoras, otra vez" vemos el cuadrado correspondiente a cada lado del triángulo rectángulo. El área del cuadrado de la hipotenusa es $\|\boldsymbol{y}\|_s^2$; el área del cuadrado del cateto que va por la recta $\mathcal{L}(\mathbf{1})$ es $\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s^2$; y el área del cuadrado del cateto perpendicular es $\sigma_{\boldsymbol{y}}^2$. El teorema de Pitágoras dice que el área grande es la suma de las dos pequeñas.

Basándonos en esta identidad, vamos a obtener una segunda expresión algebráica para calcular la varianza de un vector. Primero veamos que $\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s^2$ es simplemente $\mu_{\boldsymbol{y}}^2$. El motivo procede de la *homogeneidad de la norma* (vista en la lección 2): para cualquier escalar λ y vector $\boldsymbol{v}$, $\|\lambda\boldsymbol{v}\|_s = |\lambda| \cdot \|\boldsymbol{v}\|_s$. Aplicándolo con $\lambda = \mu_{\boldsymbol{y}}$ y $\boldsymbol{v} = \mathbf{1}$:

$$\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s = |\mu_{\boldsymbol{y}}| \cdot \|\mathbf{1}\|_s.$$

Y como fijamos en la lección 4 que $\|\mathbf{1}\|_s = 1$ (esa fue precisamente la razón de introducir el factor $\frac{1}{n}$), tenemos $\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s = |\mu_{\boldsymbol{y}}|$, y al elevar al cuadrado, $\|\mu_{\boldsymbol{y}}\mathbf{1}\|_s^2 = \mu_{\boldsymbol{y}}^2$. Sin este paso —que solo funciona porque $\|\mathbf{1}\|_s = 1$— la fórmula de cálculo no tendría esta forma tan simple.

En segundo lugar tenemos que $\|\boldsymbol{y}\|_s^2 = \langle \boldsymbol{y}, \boldsymbol{y}\rangle_s = \frac{1}{n}\sum_{i=1}^n y_i^2$; pero esta última expresión es la media aritmética del cuadrado de las componentes del vector $\boldsymbol{y}$. Si denotamos con $\boldsymbol{y}^2$ con dicho vector, obtenemos esta bonita relación $\|\boldsymbol{y}\|_s^2 = \mu_{\boldsymbol{y}^2} = \langle \boldsymbol{y}^2, \mathbf{1}\rangle_s$.

Sustituyendo en Pitágoras, $\mu_{\boldsymbol{y}^2} = \mu_{\boldsymbol{y}}^2 + \sigma_{\boldsymbol{y}}^2$, y despejando:

$$\sigma_{\boldsymbol{y}}^2 = \mu_{\boldsymbol{y}^2} - \mu_{\boldsymbol{y}}^2.$$

Comprobación con el vector de la figura de la transparencia de apertura. El vector dibujado es $\boldsymbol{y} = (1, 5)$. Su media es $\mu_{\boldsymbol{y}} = \frac{1+5}{2} = 3$. Por un lado, directamente:

$$\sigma_{\boldsymbol{y}}^2 = \frac{1}{2}[(1-3)^2 + (5-3)^2] = \frac{1}{2}(4+4) = 4.$$

Por otro, con la fórmula de cálculo:

$$\|\boldsymbol{y}\|_s^2 = \frac{1}{2}(1^2 + 5^2) = \frac{26}{2} = 13, \qquad \sigma_{\boldsymbol{y}}^2 = 13 - 3^2 = 13 - 9 = 4.$$

Ambos caminos coinciden, como debían: son la misma cosa vista desde dos ángulos distintos (uno directo, otro pasando por Pitágoras).

Este es un buen momento para insistir en la diferencia de estatus entre las dos fórmulas. La expresión $\sigma_{\boldsymbol{y}}^2 = \frac{1}{n}\sum(y_i - \mu_{\boldsymbol{y}})^2$ es la *definición*: mide directamente la dispersión respecto a la media. La fórmula $\sigma_{\boldsymbol{y}}^2 = \mu_{\boldsymbol{y}^2} - \mu_{\boldsymbol{y}}^2$ es una *consecuencia* geométrica, útil para el cálculo a mano (solo hace falta $\sum y_i$ y $\sum y_i^2$, sin calcular cada diferencia), pero que sin el contexto geométrico oculta su naturaleza. Sin embargo, con la geometría en la mano, la fórmula de cálculo deja de ser un truco de manipulación algebráica: es, sencillamente, el teorema de Pitágoras.

Para profundizar:

- Strang, G. *Introduction to Linear Algebra* (2016), sección sobre varianza y desviación típica como norma en el capítulo dedicado a estadística.

4 De un vector a dos: la covarianza

Dados $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, consideremos sus *vectores en desviaciones* (cada uno respecto de su propia media): $\boldsymbol{x} - \overline{\boldsymbol{x}}$ e $\boldsymbol{y} - \overline{\boldsymbol{y}}$.

Definición. La *covarianza* entre $\boldsymbol{x}$ e $\boldsymbol{y}$ es el producto escalar de sus vectores en desviaciones:

$$\sigma_{\boldsymbol{xy}} = \langle \boldsymbol{x} - \overline{\boldsymbol{x}},\, \boldsymbol{y} - \overline{\boldsymbol{y}} \rangle_s = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{\boldsymbol{x}})(y_i - \mu_{\boldsymbol{y}}).$$

- Si $\boldsymbol{y} = \boldsymbol{x}$: $\sigma_{\boldsymbol{xx}} = \sigma_{\boldsymbol{x}}^2$ (la varianza es la covarianza de un vector consigo mismo).
- *Simetría*: $\sigma_{\boldsymbol{xy}} = \sigma_{\boldsymbol{yx}}$.
- *Linealidad* en cada argumento.

Todo heredado del producto escalar: no hay nada nuevo que demostrar.

De un vector a dos: la covarianza

Hasta ahora trabajábamos con un único vector de datos $\boldsymbol{y}$ y su descomposición ortogonal en su vector de medias (su componente constante) y vector en desviaciones (su componente variable). La novedad de hoy es tener *dos* vectores de datos, $\boldsymbol{x}$ e $\boldsymbol{y}$, cada uno con su propia media ($\mu_{\boldsymbol{x}}$ y $\mu_{\boldsymbol{y}}$) y su propio vector en desviaciones ($\boldsymbol{x} - \overline{\boldsymbol{x}}$ y $\boldsymbol{y} - \overline{\boldsymbol{y}}$). Ambos vectores en desviaciones viven en el mismo subespacio, $\mathcal{L}(\mathbf{1})^{\perp}$ (el conjunto de todos los vectores ortogonales a $\mathbf{1}$, i.e., los vectores de media nula), así que tiene sentido preguntarse cómo se relacionan entre sí dentro de ese subespacio.

Una sencilla pregunta que nos podemos hacer sobre dos vectores es ¿cuánto vale su producto escalar? Esta pregunta, aplicada a los vectores en desviaciones, tiene nombre propio en estadística: la *covarianza*.

Definición. La covarianza entre $\boldsymbol{x}$ e $\boldsymbol{y}$ es

$$\sigma_{\boldsymbol{xy}} = \langle \boldsymbol{x} - \overline{\boldsymbol{x}},\, \boldsymbol{y} - \overline{\boldsymbol{y}} \rangle_s = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{\boldsymbol{x}})(y_i - \mu_{\boldsymbol{y}}).$$

Un caso particular: si tomamos $\boldsymbol{y} = \boldsymbol{x}$, la covarianza se convierte en $\sigma_{\boldsymbol{xx}} = \langle \boldsymbol{x} - \overline{\boldsymbol{x}}, \boldsymbol{x} - \overline{\boldsymbol{x}} \rangle_s = \|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_s^2 = \sigma_{\boldsymbol{x}}^2$. Es decir, *la varianza es la covarianza de un vector consigo mismo.* Esto no es una coincidencia: la varianza es, geométricamente, el caso degenerado de la covarianza. Consecuentemente, todo lo que digamos sobre la covarianza se aplicará, como caso particular, a la varianza.

Como la covarianza es (por definición) un producto escalar, hereda de manera automática sus propiedades: es *simétrica* ($\sigma_{\boldsymbol{xy}} = \sigma_{\boldsymbol{yx}}$, porque $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_s = \langle \boldsymbol{b}, \boldsymbol{a} \rangle_s$) y *lineal* en cada uno de sus dos argumentos (con respecto al primer argumento: $\sigma_{(\alpha\boldsymbol{x}+\beta\boldsymbol{y})\boldsymbol{z}} = \alpha\sigma_{\boldsymbol{xz}} + \beta\sigma_{\boldsymbol{yz}}$; y de manera similar respecto al segundo). No hace falta demostrar nada nuevo: son las mismas propiedades del producto escalar que fijamos en la lección 2, aplicadas ahora a vectores en desviaciones.

Una advertencia importante, que retomaremos en la siguiente transparencia: la covarianza, por sí sola, no tiene una escala interpretable. Su valor numérico depende de las unidades en que midamos $\boldsymbol{x}$ e $\boldsymbol{y}$ (si $\boldsymbol{x}$ está en euros y lo pasamos a céntimos, la covarianza se multiplica por 100, sin que la

naturaleza de la posible relación entre ambas variables haya cambiado en absoluto, pues solo hemos cambiado la unidad de medida). Esto es lo mismo que le ocurre al producto escalar ordinario: por sí solo no nos dice nada sobre el *ángulo* entre dos vectores, salvo que lo dividamos por sus normas; lo que produce un ratio sin unidades de medida (pues las unidades del numerador y del denominador se cancelan). Ese es el papel que va a jugar la correlación.

5 La correlación: el coseno de un ángulo

Definición. La *correlación* entre dos vectores no constantes $\boldsymbol{x}$ e $\boldsymbol{y}$ es el coseno del ángulo θ que forman sus vectores en desviaciones:

$$\rho_{\boldsymbol{xy}} = \cos\theta = \frac{\langle \boldsymbol{x}-\overline{\boldsymbol{x}},\, \boldsymbol{y}-\overline{\boldsymbol{y}}\rangle_s}{\|\boldsymbol{x}-\overline{\boldsymbol{x}}\|_s\, \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}\,\sigma_{\boldsymbol{y}}}.$$

Por Cauchy–Schwarz (lección 3): $-1 \leq \rho_{\boldsymbol{xy}} \leq 1$.

- $\rho_{\boldsymbol{xy}} = \pm 1$: vectores en desviaciones *alineados* ($\theta = 0^o$ o 180^o).
- $\rho_{\boldsymbol{xy}} = 0$: vectores en desviaciones *ortogonales* ($\theta = 90^o$).

A diferencia de la covarianza, la correlación no depende de las unidades: es algún número en el intervalo $[-1, 1]$.

La correlación: el coseno de un ángulo

En la lección 3 definimos el coseno del ángulo entre dos vectores cualesquiera como $\cos\theta = \frac{\langle \boldsymbol{a},\boldsymbol{b}\rangle}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}$, y demostramos que la desigualdad de Cauchy–Schwarz garantiza que ese cociente siempre cae en el intervalo de valores $[-1, 1]$. No hay nada que impida aplicar la misma definición a los vectores en desviaciones $\boldsymbol{x}-\overline{\boldsymbol{x}}$ e $\boldsymbol{y}-\overline{\boldsymbol{y}}$: ambos son vectores de $\mathbb{R}^n$ como cualquier otro, y todo lo demostrado en la lección 3 (Pitágoras, Cauchy–Schwarz, la desigualdad triangular) sigue siendo válido, porque se apoya únicamente en las propiedades abstractas del producto escalar, no en cuáles sean concretamente los vectores[1].

Definición. La correlación entre dos vectores no constantes $\boldsymbol{x}$ e $\boldsymbol{y}$ es

$$\rho_{\boldsymbol{xy}} = \frac{\langle \boldsymbol{x}-\overline{\boldsymbol{x}},\, \boldsymbol{y}-\overline{\boldsymbol{y}}\rangle_s}{\|\boldsymbol{x}-\overline{\boldsymbol{x}}\|_s\, \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}}.$$

Es, literalmente, el coseno del ángulo θ que forman los dos vectores en desviaciones. La desigualdad de Cauchy–Schwarz —demostrada con detalle en la lección 3— nos garantiza que $-1 \leq \rho_{\boldsymbol{xy}} \leq 1$ siempre, para cualesquiera $\boldsymbol{x}$ e $\boldsymbol{y}$.

Los dos casos extremos son especialmente informativos:

- $\rho_{\boldsymbol{xy}} = \pm 1$ ocurre exactamente cuando los vectores en desviaciones están *alineados* (uno es múltiplo escalar no nulo del otro): recordemos que en la lección 3 vimos que la igualdad en Cauchy–Schwarz se da si y solo si los vectores son proporcionales.

[1] Fíjese que la definición de la correlación emplea específicamente el producto escalar de la estadística

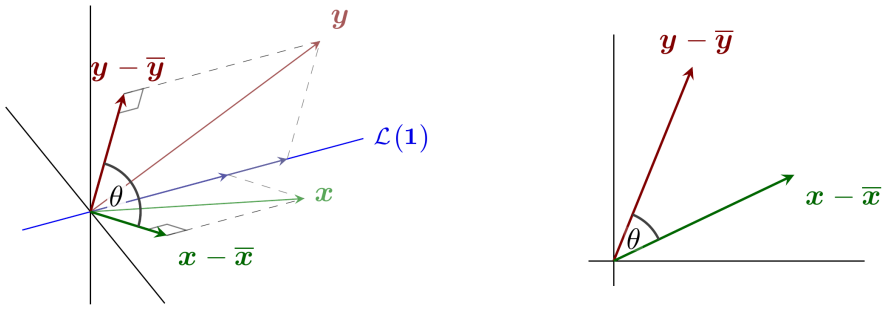
- $\rho_{\boldsymbol{xy}} = 0$ ocurre exactamente cuando los vectores en desviaciones son *ortogonales*. Este es el sentido geométrico preciso de "no estar correlacionados": no que $\boldsymbol{x}$ e $\boldsymbol{y}$ no tengan relación alguna, sino que sus desviaciones respecto a sus medias son perpendiculares.

Nótese la diferencia de estatus con la covarianza. La covarianza cambia de valor si cambiamos las unidades de medida (multiplicar $\boldsymbol{x}$ por una constante multiplica $\sigma_{\boldsymbol{xy}}$ por esa misma constante; propiedad de los productos escalares que revisitaremos en unas transparencias). La correlación, al ser un coseno, es adimensional: no depende de las unidades. Esta es la razón por la que la correlación —y no la covarianza— es la medida habitual para comparar la fuerza de la relación *lineal* entre las componentes no constantes de dos variables.

Para profundizar:

- Wooldridge, J. *Introductory Econometrics* (2020), Apéndice C.3: covarianza y correlación como medidas de asociación lineal.

6 Ilustración geométrica de la correlación



Dos vectores $\boldsymbol{y}$ (rojo) y $\boldsymbol{x}$ (verde) en $\mathbb{R}^3$, con sus vectores en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\boldsymbol{x} - \overline{\boldsymbol{x}}$ en el plano perpendicular al eje azul $\mathcal{L}(\mathbf{1})$. El *coseno del ángulo* θ entre esos dos vectores en desviaciones es la correlación $\rho_{\boldsymbol{xy}}$. La figura de la derecha es la misma escena, contemplada desde un punto de vista distinto (perpendicular a ambos vectores en desviaciones para apreciar mejor el ángulo θ). (versión interactiva, parte 2)

La correlación indica el "grado de alineación" entre las componentes no constantes.

Ilustración geométrica de la correlación

La figura retoma la escena de la lección 4 (dos vectores en $\mathbb{R}^3$ proyectados sobre $\mathcal{L}(\mathbf{1})$), pero añade un elemento nuevo: el ángulo θ entre los vectores en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\boldsymbol{x} - \overline{\boldsymbol{x}}$, ambos en el plano $\mathcal{L}(\mathbf{1})^{\perp}$ (perpendicular al eje azul). El coseno de dicho ángulo es la correlación $\rho_{\boldsymbol{xy}}$.

La primera imagen muestra la escena desde un punto de vista arbitrario de $\mathbb{R}^3$; en ella el ángulo θ es difícil de apreciar a simple vista, porque los dos vectores en desviaciones no están en el plano de la página (o pantalla del proyector). La segunda imagen es la misma escena, pero vista desde un punto de observación situado sobre la propia dirección ortogonal a ambos vectores en desviaciones: desde ahí, el plano $\mathcal{L}(\mathbf{1})^{\perp}$ se ve "de frente", y el ángulo θ se aprecia con su verdadera magnitud.

Una observación: para que dos vectores en desviaciones puedan formar un ángulo distinto de 0^o o 180^o, necesitamos que $\mathcal{L}(\mathbf{1})^\perp$ tenga al menos dos dimensiones, es decir, necesitamos una cantidad de datos $n \geq 3$ en cada vector. La siguiente transparencia lo explica.

7 Un hecho geométrico: en $\mathbb{R}^2$ la correlación es siempre ± 1

En $\mathbb{R}^2$, $\mathcal{L}(\mathbf{1})^\perp$ tiene dimensión $2-1=1$: es una *recta*.

Todo vector en desviaciones en $\mathbb{R}^2$ vive en *esa misma recta*.

Por tanto, dos vectores en desviaciones cualesquiera en $\mathbb{R}^2$ están *siempre alineados*:

$$\rho_{\boldsymbol{xy}} = \pm 1 \qquad (\text{en } \mathbb{R}^2).$$

Comprobación, con los ejemplos de la lección 4: $\boldsymbol{x} = (2, 4)$ y $\boldsymbol{z} = (2, -6)$ dan $\rho_{\boldsymbol{xz}} = -1$.

Para ver ángulos *genuinos* hace falta $n \geq 3$. Por eso las figuras de hoy viven en $\mathbb{R}^3$.

Un hecho geométrico: en $\mathbb{R}^2$ la correlación es siempre ± 1

Aquí hay una consecuencia geométrica que conviene entender bien, pues explica por qué las figuras de esta lección requieren necesariamente escenas en $\mathbb{R}^3$.

El subespacio $\mathcal{L}(\mathbf{1})^\perp$ —el conjunto de todos los vectores ortogonales a $\mathbf{1}$, es decir, el conjunto de todos los posibles vectores en desviaciones— tiene dimensión $n-1$ dentro de $\mathbb{R}^n$ (porque $\mathcal{L}(\mathbf{1})$, al ser una recta, ya "ocupa" una dimensión).[2] En el espacio bidimensional $\mathbb{R}^2$, eso significa que $\mathcal{L}(\mathbf{1})^\perp$ tiene dimensión $2-1=1$: es una simple *recta* que pasa por el origen, perpendicular a $\mathbf{1}$.

Ahora bien: en una recta, *todos* los vectores no nulos son múltiplos escalares unos de otros (dos vectores en una recta están, por definición, alineados). Así que, para cualesquiera dos vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ (ninguno de ellos constante), sus vectores en desviaciones $\boldsymbol{x} - \overline{\boldsymbol{x}}$ e $\boldsymbol{y} - \overline{\boldsymbol{y}}$ viven forzosamente en esa misma recta $\mathcal{L}(\mathbf{1})^\perp$, y por tanto están *siempre alineados*. Recordando la lección 3: vectores alineados dan $\cos\theta = \pm 1$. Es decir:

> Con solo dos observaciones ($n = 2$), la correlación entre dos variables cualesquiera es siempre $\rho = +1$ o $\rho = -1$ (salvo que alguna tenga varianza cero, en cuyo caso la correlación ni siquiera está definida).

Verifiquémoslo con dos de los ejemplos numéricos de la lección 4. Tomamos $\boldsymbol{x} = (2, 4)$ (con $\mu_{\boldsymbol{x}} = 3$, en desviaciones $(-1, 1)$ y $\sigma_{\boldsymbol{x}} = 1$) y $\boldsymbol{z} = (2, -6)$ (con $\mu_{\boldsymbol{z}} = -2$, en desviaciones $(4, -4)$ y $\sigma_{\boldsymbol{z}} = 4$). La covarianza es

$$\sigma_{\boldsymbol{xz}} = \frac{1}{2}\big[(-1)(4) + (1)(-4)\big] = \frac{1}{2}(-8) = -4.$$

[2] *Nota sobre la palabra "dimensión".* En este curso no hemos definido axiomáticamente qué es la dimensión de un subespacio (evitamos deliberadamente la maquinaria abstracta del álgebra lineal). La usamos aquí en su sentido más intuitivo: el número de "direcciones independientes" que caben dentro de un subespacio, la misma idea que ya manejábamos visualmente al hablar de rectas (una dirección) y planos (dos direcciones). Esta noción informal es todo lo que necesitamos para lo que sigue.

Y la correlación,

$$\rho_{\boldsymbol{xz}} = \frac{-4}{1 \cdot 4} = -1.$$

Efectivamente, -1: los vectores en desviaciones $(-1, 1)$ y $(4, -4)$ son múltiplos el uno del otro (de hecho, $(4, -4) = -4 \cdot (-1, 1)$), así que están alineados, pero con sentidos opuestos.

Este hecho no es una peculiaridad de estos dos vectores concretos: es forzoso en $\mathbb{R}^2$, sea cual sea el par de vectores que elijamos. La consecuencia práctica es importante: para ilustrar con un dibujo una correlación *distinta* de ± 1 (es decir, un ángulo θ genuino, ni nulo ni llano) necesitamos, como mínimo, $\mathbb{R}^3$ ($n = 3$ observaciones), porque solo entonces $\mathcal{L}(\mathbf{1})^\perp$ tiene dimensión 2 y puede albergar de verdad dos direcciones distintas. Por eso las figuras de esta lección —y el ejemplo numérico que viene a continuación— se sitúan en $\mathbb{R}^3$.

Para profundizar:

- Strang, G. *Introduction to Linear Algebra* (2016), capítulo 4: dimensión de subespacios y su relación con el número de direcciones independientes (tratamiento formal de la noción que aquí usamos de manera informal).

8 Propiedades geométricas: traslación y escala

Invarianza ante traslaciones sumar una constante a $\boldsymbol{x}$ o a $\boldsymbol{y}$ *no* cambia $\sigma_{\boldsymbol{xy}}$ ni $\rho_{\boldsymbol{xy}}$.

Homogeneidad ante escala si $\boldsymbol{x}' = \lambda \boldsymbol{x}$ con $\lambda \neq 0$:

$$\sigma_{\boldsymbol{x'y}} = \lambda\, \sigma_{\boldsymbol{xy}}, \qquad \rho_{\boldsymbol{x'y}} = \frac{\lambda}{|\lambda|}\, \rho_{\boldsymbol{xy}} = \pm \rho_{\boldsymbol{xy}}.$$

El caso $\boldsymbol{y} = \boldsymbol{x}$ (la varianza):

$$\sigma^2_{\boldsymbol{x}+a\mathbf{1}} = \sigma^2_{\boldsymbol{x}}, \qquad \sigma^2_{\lambda \boldsymbol{x}} = \lambda^2\, \sigma^2_{\boldsymbol{x}}.$$

La correlación es invariante a cambios de escala (salvo, quizá, el signo); pero la covarianza no.

Por otra parte, la desviación típica se escala por $|\lambda|$ (y la varianza, por λ^2).

Propiedades geométricas: traslación y escala

Estas dos propiedades son la versión, para dos vectores, de las que ya vimos en la lección 4 para la desviación típica de un solo vector.

Invarianza ante traslaciones. Si $\boldsymbol{x}' = \boldsymbol{x} + a\mathbf{1}$ (sumamos una constante a a todas las componentes de $\boldsymbol{x}$), entonces $\mu_{\boldsymbol{x}'} = \mu_{\boldsymbol{x}} + a$, pero el vector en desviaciones (la *componente no constante*) no cambia:

$$\boldsymbol{x}' - \overline{\boldsymbol{x}'} = (\boldsymbol{x} + a\mathbf{1}) - (\mu_{\boldsymbol{x}} + a)\mathbf{1} = \boldsymbol{x} - \overline{\boldsymbol{x}}.$$

Como la covarianza y la correlación solo dependen de los vectores en desviaciones, ni $\sigma_{\boldsymbol{x'y}}$ ni $\rho_{\boldsymbol{x'y}}$ cambian. Geométricamente: sumar $a\mathbf{1}$ desplaza el vector a lo largo de $\mathcal{L}(\mathbf{1})$, sin tocar su componente en $\mathcal{L}(\mathbf{1})^\perp$.

Homogeneidad ante escala. Si $\boldsymbol{x'} = \lambda \boldsymbol{x}$ con $\lambda \neq 0$, entonces $\mu_{\boldsymbol{x'}} = \lambda \mu_{\boldsymbol{x}}$; pero ahora, además, el vector en desviaciones también se escala por λ:

$$\boldsymbol{x'} - \overline{\boldsymbol{x'}} = \lambda \boldsymbol{x} - \lambda \mu_{\boldsymbol{x}} \mathbf{1} = \lambda (\boldsymbol{x} - \overline{\boldsymbol{x}}).$$

Por la linealidad del producto escalar,

$$\sigma_{\boldsymbol{x'y}} = \langle \lambda(\boldsymbol{x} - \overline{\boldsymbol{x}}),\, \boldsymbol{y} - \overline{\boldsymbol{y}} \rangle_s = \lambda\, \sigma_{\boldsymbol{xy}}.$$

Y como $\|\lambda(\boldsymbol{x} - \overline{\boldsymbol{x}})\|_s = |\lambda|\, \|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_s$ (homogeneidad de la norma, lección 2), la correlación queda afectada del siguiente modo:

$$\rho_{\boldsymbol{x'y}} = \frac{\lambda\, \sigma_{\boldsymbol{xy}}}{|\lambda|\, \sigma_{\boldsymbol{x}} \sigma_{\boldsymbol{y}}} = \frac{\lambda}{|\lambda|}\, \rho_{\boldsymbol{xy}}.$$

Si $\lambda > 0$, $\frac{\lambda}{|\lambda|} = 1$ y la correlación no cambia. Si $\lambda < 0$, $\frac{\lambda}{|\lambda|} = -1$ y la correlación cambia de signo, pero conserva su magnitud $|\rho_{\boldsymbol{xy}}| = |\rho_{\boldsymbol{x'y}}|$.

Caso particular: la varianza. Tomando $\boldsymbol{y} = \boldsymbol{x}$ en las dos propiedades anteriores obtenemos, de forma inmediata, las propiedades correspondientes de la varianza —el caso degenerado $\sigma_{\boldsymbol{xx}} = \sigma^2_{\boldsymbol{x}}$ que ya vimos en la transparencia de la covarianza—. De la invarianza ante traslaciones:

$$\sigma^2_{\boldsymbol{x}+a\mathbf{1}} = \sigma_{(\boldsymbol{x}+a\mathbf{1})(\boldsymbol{x}+a\mathbf{1})} = \sigma_{\boldsymbol{xx}} = \sigma^2_{\boldsymbol{x}}.$$

Y de la homogeneidad ante escala, aplicada dos veces —una por cada argumento de $\sigma_{\boldsymbol{xx}}$, ya que ambos son $\boldsymbol{x}$—:

$$\sigma^2_{\lambda\boldsymbol{x}} = \sigma_{(\lambda\boldsymbol{x})(\lambda\boldsymbol{x})} = \lambda\, \sigma_{\boldsymbol{x}(\lambda\boldsymbol{x})} = \lambda \cdot \lambda\, \sigma_{\boldsymbol{xx}} = \lambda^2\, \sigma^2_{\boldsymbol{x}}.$$

Otra forma más directa de verlo: esto es también consecuencia inmediata de la homogeneidad de la norma (lección 2): $\|\lambda(\boldsymbol{x} - \overline{\boldsymbol{x}})\|_s = |\lambda|\, \|\boldsymbol{x} - \overline{\boldsymbol{x}}\|_s$; por lo que al elevar al cuadrado la desviación típica (que es una norma), tenemos que $|\lambda|^2 = \lambda^2$.

Ambas identidades ya estaban, en realidad, implícitas en la lección 4 (sección "Propiedades geométricas"): allí vimos que sumar una constante no cambia la desviación típica $\sigma_{\boldsymbol{x}}$, lo cual —al elevar al cuadrado— es exactamente la invarianza de la varianza ante traslaciones. Lo que añadimos aquí es el efecto, *cuadrático*, de la escala: si $\boldsymbol{x}$ pasa de medirse en euros a medirse en céntimos ($\lambda = 100$), la desviación típica se multiplica por 100, pero la varianza se multiplica por $100^2 = 10\,000$, no por 100. Confundir ambos efectos es una fuente frecuente de errores al comparar varianzas expresadas en unidades distintas.

La lectura práctica: la covarianza cambia con las unidades (si $\boldsymbol{x}$ pasa de metros a centímetros, $\sigma_{\boldsymbol{xy}}$ se multiplica por 100), pero la correlación —al ser un coseno— es invariante a cambios de escala (salvo, cuando la escala es negativa, un cambio de signo). Esta es la razón geométrica de por qué la correlación, y no la covarianza, es la medida de asociación lineal que se puede comparar entre variables medidas en unidades distintas.

Para profundizar:

- Wooldridge, J. *Introductory Econometrics* (2020), Apéndice C: efecto de las transformaciones lineales sobre la media, la varianza y la covarianza.

9 Ejemplo numérico completo en $\mathbb{R}^3$

	Primer Vector $\boldsymbol{x}$	Segundo Vector $\boldsymbol{y}$
Datos	$\boldsymbol{x} = (1, 2, 3)$	$\boldsymbol{y} = (2, 2, 5)$
Media	$\mu_{\boldsymbol{x}} = 2$	$\mu_{\boldsymbol{y}} = 3$
Vector en desviaciones	$(-1, 0, 1)$	$(-1, -1, 2)$
Varianza	$\sigma^2_{\boldsymbol{x}} = \frac{2}{3}$	$\sigma^2_{\boldsymbol{y}} = 2$

Covarianza : $\sigma_{\boldsymbol{xy}} = \frac{1}{3}[(-1)(-1) + (0)(-1) + (1)(2)] = 1.$

$$\rho_{\boldsymbol{xy}} = \frac{1}{\sqrt{\frac{2}{3} \cdot 2}} = \frac{1}{\sqrt{4/3}} = \frac{\sqrt{3}}{2} \approx 0{,}866.$$

Un ángulo $\theta = 30^o$: ni ortogonales, ni alineados. Un caso *imposible* en $\mathbb{R}^2$.

Ejemplo numérico completo

Verifiquemos todo lo dicho con un ejemplo completo. Necesitamos $n \geq 3$ para que la correlación pueda tomar un valor distinto de ± 1, así que trabajamos en $\mathbb{R}^3$: $\boldsymbol{x} = (1, 2, 3)$, $\boldsymbol{y} = (2, 2, 5)$.

Medias:

$$\mu_{\boldsymbol{x}} = \frac{1 + 2 + 3}{3} = 2, \qquad \mu_{\boldsymbol{y}} = \frac{2 + 2 + 5}{3} = 3.$$

Vectores en desviaciones:

$$\boldsymbol{x} - \overline{\boldsymbol{x}} = (1 - 2,\, 2 - 2,\, 3 - 2) = (-1, 0, 1), \qquad \boldsymbol{y} - \overline{\boldsymbol{y}} = (2 - 3,\, 2 - 3,\, 5 - 3) = (-1, -1, 2).$$

Covarianza:

$$\sigma_{\boldsymbol{xy}} = \frac{1}{3}[(-1)(-1) + (0)(-1) + (1)(2)] = \frac{1}{3}(1 + 0 + 2) = 1.$$

Varianzas:

$$\sigma^2_{\boldsymbol{x}} = \frac{1}{3}[(-1)^2 + 0^2 + 1^2] = \frac{2}{3}, \qquad \sigma^2_{\boldsymbol{y}} = \frac{1}{3}[(-1)^2 + (-1)^2 + 2^2] = \frac{6}{3} = 2.$$

Correlación:

$$\rho_{\boldsymbol{xy}} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}} = \frac{1}{\sqrt{\frac{2}{3}} \cdot \sqrt{2}} = \frac{1}{\sqrt{\frac{4}{3}}} = \frac{\sqrt{3}}{2} \approx 0{,}866.$$

Comprobamos que se cumple Cauchy–Schwarz con desigualdad *estricta*: $\sigma^2_{\boldsymbol{xy}} = 1 \leq \sigma^2_{\boldsymbol{x}}\sigma^2_{\boldsymbol{y}} = \frac{2}{3} \cdot 2 = \frac{4}{3}$, y en efecto $1 < \frac{4}{3}$, señal de que los vectores en desviaciones *no* están alineados (si lo estuvieran, tendríamos igualdad, y $\rho = \pm 1$).

El valor $\rho_{\boldsymbol{xy}} = \frac{\sqrt{3}}{2}$ corresponde a un ángulo $\theta = 30^o$ entre los vectores en desviaciones $(-1, 0, 1)$ y $(-1, -1, 2)$: ni ortogonales ($\theta \neq 90^o$, $\rho \neq 0$), ni alineados ($\theta \neq 0^o, 180^o$, $\rho \neq \pm 1$). Es exactamente el tipo de situación que, como vimos al estudiar la correlación en $\mathbb{R}^2$, es *imposible* de obtener con solo $n = 2$ observaciones (por eso hemos acudido a un ejemplo en $\mathbb{R}^3$).

10 Recapitulación

Objeto/Concepto	Definición geométrica/algebraica	Fórmula en estadística descriptiva
Varianza $\sigma_{\boldsymbol{x}}^2$	$\|\boldsymbol{x}-\overline{\boldsymbol{x}}\|_s^2$	$\frac{1}{n}\sum(x_i-\mu_{\boldsymbol{x}})^2 = \mu_{\boldsymbol{x}^2}-\mu_{\boldsymbol{x}}^2$
Covarianza $\sigma_{\boldsymbol{xy}}$	$\langle \boldsymbol{x}-\overline{\boldsymbol{x}},\, \boldsymbol{y}-\overline{\boldsymbol{y}}\rangle_s$	$\frac{1}{n}\sum(x_i-\mu_{\boldsymbol{x}})(y_i-\mu_{\boldsymbol{y}})$
Correlación $\rho_{\boldsymbol{xy}}$	cos del ángulo θ entre $(\boldsymbol{x}-\overline{\boldsymbol{x}})$ e $(\boldsymbol{y}-\overline{\boldsymbol{y}})$	$\dfrac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}}$

El principio unificador: la covarianza es un producto escalar; la correlación, un coseno. Nada de esto es nuevo: es la geometría de las lecciones 2 y 3, aplicada ahora a los vectores en desviaciones.

Recapitulación

El mensaje central de esta sesión es que, al pasar de un vector a dos, no necesitamos ningún concepto nuevo: la covarianza y la correlación son, respectivamente, el producto escalar y el coseno del ángulo, aplicados a los vectores en desviaciones. Todo lo que sabíamos sobre productos escalares y ángulos (lecciones 2 y 3) se traslada sin cambios.

Tres consecuencias prácticas:

1. La varianza es el caso particular $\sigma_{\boldsymbol{xx}} = \sigma_{\boldsymbol{x}}^2$ de la covarianza.
2. La desigualdad de Cauchy–Schwarz, demostrada en la lección 3, es la que garantiza $-1 \leq \rho_{\boldsymbol{xy}} \leq 1$: no hace falta ninguna demostración nueva para justificar por qué la correlación nunca se sale de ese intervalo.
3. La dimensión de $\mathcal{L}(\mathbf{1})^\perp$ importa: en $\mathbb{R}^2$ es una recta, y por eso la correlación entre cualquier par de vectores en $\mathbb{R}^2$ es forzosamente ± 1. Se necesita $n \geq 3$ para observar correlaciones intermedias.

La lección 6 llevará este mismo aparato —proyecciones, ortogonalidad, ángulos— a un problema distinto: buscaremos la combinación lineal de $\mathbf{1}$ y $\boldsymbol{x}$ que mejor aproxima a $\boldsymbol{y}$. Ese será el problema de la regresión lineal simple.

11 Guiño a la sesión siguiente

Próxima sesión: laboratorio en Gretl. Estadísticos descriptivos, correlaciones y diagramas de dispersión con datos reales; verificación numérica de las identidades de hoy (Pitágoras, covarianza, correlación) con datos simulados.

Lección 6: la regresión lineal simple. Añadiremos un segundo regresor $\boldsymbol{x}$ a la proyección sobre $\mathcal{L}(\mathbf{1})$: proyectaremos sobre $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, ya no una recta sino un *plano*. El ángulo θ de hoy reaparecerá en la lección 8 como $R^2 = \cos^2\theta$.

Guiño a la sesión siguiente

Hoy hemos visto que para relacionar dos vectores no se exige ningún concepto matemáticamente nuevo: la covarianza es un producto escalar (entre vectores en desviaciones) y la correlación es un coseno (entre esos mismos vectores en desviaciones). La próxima sesión es un laboratorio: en

él calcularemos, con datos reales, medias, varianzas, covarianzas y correlaciones, y dibujaremos diagramas de dispersión; y comprobaremos, con datos simulados, que las identidades de hoy (en particular, la fórmula de cálculo de la varianza y la relación $\rho = \cos\theta$) se cumplen exactamente.

La lección 6 dará el siguiente gran paso del curso: la *regresión lineal simple*. Hasta ahora hemos proyectado un único vector $\boldsymbol{y}$ sobre la recta $\mathcal{L}(\mathbf{1})$ (lección 4); ahora añadiremos un segundo regresor $\boldsymbol{x}$ y proyectaremos sobre el *plano* $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ (el conjunto de todas las combinaciones lineales de $\mathbf{1}$ y $\boldsymbol{x}$). La lógica —imponer condiciones de ortogonalidad— será la misma que ya conocemos; solo cambia el subespacio sobre el que proyectamos. Y al final de ese camino, cuando preguntemos "¿cómo de bueno es el ajuste?", la respuesta será, el coseno al cuadrado de un ángulo: $R^2 = \cos^2\theta$, donde θ es el ángulo entre el vector en desviaciones y su ajuste en desviaciones.[3] El ángulo de hoy no desaparece: reaparece con otro nombre.

Por último, una pincelada de lo que vendrá más lejos: cuando haya varios regresores, resultará cómodo agruparlos como columnas de una matriz. Esa notación llegará en la lección 10, cuando de verdad la necesitemos; y allí veremos que la doble lectura de la media de hoy —producto escalar $\langle \boldsymbol{y}, \mathbf{1} \rangle_s$ y, a la vez, coeficiente de la proyección $\mu_{\boldsymbol{y}}\mathbf{1}$— reaparece con varios regresores.

[3] recuerde que *vector en desviaciones* respecto a la media y *vector centrado* son sinónimos.