Índice general

Lección 8. Bondad de ajuste: de Pitágoras a R^2

Marcos Bujosa

19 de septiembre de 2026

Resumen

Las dos ortogonalidades de la lección 7, vistas bajo Pitágoras, descomponen la varianza de $\boldsymbol{y}$ en la del ajuste más la del error. El cociente es el cuadrado del coseno del ángulo entre datos y ajuste, ambos en desviaciones: R^2.

"Un triángulo rectángulo escondido en cada regresión."

- (slides) — (html) — (pdf)

1 De dónde venimos: dos ortogonalidades, una identidad guardada

En la lección 7 resolvimos las dos condiciones de ortogonalidad de la *regresión lineal simple*:

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

Que nos condujeron al hallazgo de dos identidades. Por una parte:

$$\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}.$$

Y por otra:

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

Hoy juntamos estas piezas con una herramienta conocida: el *teorema de Pitágoras* (lección 3).

De dónde venimos: dos ortogonalidades, una identidad guardada

La lección 7 resolvió el sistema de dos ortogonalidades que la lección 6 había dejado planteado: impusimos $\langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0$ y $\langle \widehat{\boldsymbol{e}}, \boldsymbol{x} \rangle_s = 0$, y de ahí salieron $\hat{\beta}_1 = \mu_{\boldsymbol{y}} - \hat{\beta}_2 \mu_{\boldsymbol{x}}$ y $\hat{\beta}_2 = \sigma_{\boldsymbol{xy}}/\sigma^2_{\boldsymbol{x}}$.

De la primera condición obtuvimos, además, dos consecuencias que hoy son protagonistas. Primero, que la media del vector ajustado coincide con la media de los datos: $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$ (porque $\widehat{\boldsymbol{e}} \perp \mathbf{1}$ implica $\mu_{\widehat{\boldsymbol{e}}} = 0$, y $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$). Segundo, la identidad

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

que decía, literalmente, que el vector en desviaciones del ajuste *es* el mismo vector que el vector en desviaciones de $\hat{\beta}_2\boldsymbol{x}$. En la lección 7 la usamos únicamente para concluir que $\boldsymbol{y}$ forma el mismo

ángulo con $\widehat{\boldsymbol{y}}$ que con $\hat{\beta}_2\boldsymbol{x}$. Al final de la lección volveremos sobre esta identidad para deducir nuevas propiedades del ajuste lineal simple.

El plan de la sesión es el siguiente. Vamos a construir, a partir de $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$, un triángulo rectángulo cuya hipotenusa sea el vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$. Aplicando el teorema de Pitágoras (lección 3) a ese triángulo obtendremos una descomposición de la varianza de $\boldsymbol{y}$; de ahí surgirá, de manera casi inmediata, el coeficiente de determinación R^2 como el coseno al cuadrado de un ángulo. Y, al final, la identidad guardada nos permitirá relacionar ese R^2 con la correlación $\rho_{\boldsymbol{xy}}$ de la lección 5, cerrando un círculo que empezamos a dibujar hace varias sesiones.

Para profundizar:

- Wooldridge, J. M. (2020). *Introductory Econometrics*, cap. 2, sección 2.3: "Units of Measurement and Functional Form" y la presentación clásica de SST = SSE + SSR.

2 El triángulo rectángulo de la bondad de ajuste

Restando $\overline{\boldsymbol{y}}$ a ambos lados de $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$ tenemos:

$$\boldsymbol{y} - \overline{\boldsymbol{y}} \;=\; (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}}, \qquad \text{con } \widehat{\boldsymbol{e}} \perp \mathcal{L}(\widehat{\boldsymbol{y}}, \overline{\boldsymbol{y}}) \text{ así,}$$

$(\boldsymbol{y} - \overline{\boldsymbol{y}})$ es la *hipotenusa* de un triángulo rectángulo cuyos catetos son $\widehat{\boldsymbol{e}}$ y $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$.

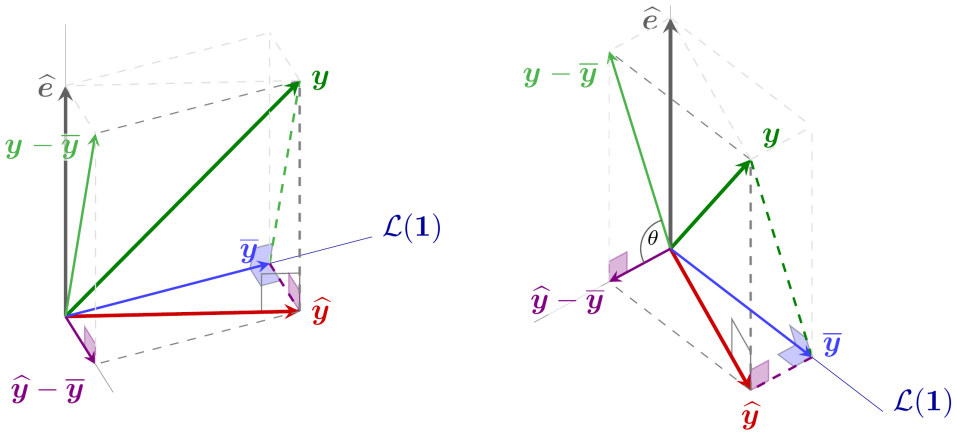

Figura 1: Proyección ortogonal de $\boldsymbol{y}$ sobre el subespacio generado por los regresores, vista desde dos ángulos. En verde oscuro el vector de datos $\boldsymbol{y}$. En azul, el vector de medias $\overline{\boldsymbol{y}}$ (proyección tanto de $\boldsymbol{y}$ como de $\widehat{\boldsymbol{y}}$ sobre $\mathcal{L}(\mathbf{1})$). En rojo, el ajuste $\widehat{\boldsymbol{y}}$ (en el plano); en violeta, su vector en desviaciones $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$. En verde claro, el vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$. En gris, el vector de residuos $\widehat{\boldsymbol{e}}$. El arco señala el ángulo θ entre los dos vectores en desviaciones. Los pequeños cuadrados señalan los ángulos rectos correspondientes a distintos triángulos rectángulos.

Versión interactiva: cuaderno 3 en MyBinder, parte 1 (el triángulo, rotable).

El triángulo rectángulo de la bondad de ajuste

Detengámonos en justificar, una por una, las cuatro marcas de ortogonalidad de la figura, porque de la última de ellas sale todo lo que sigue.

Marca 1 y *marca 2* (adyacentes a $\overline{\boldsymbol{y}}$, azules y pequeñas): Ambas conocidas desde la lección 4. Señalan dos proyecciones ortogonales sobre $\mathcal{L}(\mathbf{1})$ que coinciden en un único vector $\overline{\boldsymbol{y}}$ (múltiplo de $\mathbf{1}$). Como ya sabíamos, tanto $\boldsymbol{y}$ como $\widehat{\boldsymbol{y}}$ comparten la misma proyección $\overline{\boldsymbol{y}}$ sobre $\mathcal{L}(\mathbf{1})$ (consecuentemente $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$, lección 7).

Marca 3 (transparente y grande): $\widehat{\boldsymbol{e}} \perp \widehat{\boldsymbol{y}}$. Esta es la ortogonalidad "principal": por construcción $\widehat{\boldsymbol{e}}$ es ortogonal a *todo* el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$; y $\widehat{\boldsymbol{y}}$ está en ese plano (por ser combinación lineal de $\mathbf{1}$ y de $\boldsymbol{x}$; pues $\widehat{\boldsymbol{y}} = \hat{\beta}_1\mathbf{1} + \hat{\beta}_2\boldsymbol{x}$).

Marca 4 (**la que más nos interesa hoy**, morada, pequeña y duplicada): $\widehat{\boldsymbol{e}} \perp (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$. Como en la marca anterior, se deduce porque $\widehat{\boldsymbol{y}}$ e $\overline{\boldsymbol{y}}$ son vectores del plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$.

$$\langle \widehat{\boldsymbol{e}}, \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \rangle_s = \langle \widehat{\boldsymbol{e}}, \widehat{\boldsymbol{y}} \rangle_s - \mu_{\widehat{\boldsymbol{y}}} \langle \widehat{\boldsymbol{e}}, \mathbf{1} \rangle_s = 0 - \mu_{\widehat{\boldsymbol{y}}} \cdot 0 = 0.$$

Con las cuatro marcas confirmadas, podemos leer la descomposición central de la sesión: restando $\overline{\boldsymbol{y}}$ a ambos lados de $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$,

$$\boldsymbol{y} - \overline{\boldsymbol{y}} = (\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}) - \overline{\boldsymbol{y}} = (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}},$$

Tal como indica la marca 4 (duplicada en dos planos paralelos), los dos sumandos $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$ y $\widehat{\boldsymbol{e}}$ son ortogonales entre sí. Tenemos, por tanto, un triángulo rectángulo genuino: hipotenusa $\boldsymbol{y} - \overline{\boldsymbol{y}}$, catetos $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{e}}$.

Respecto a la figura: las dos vistas muestran la misma escena en $\mathbb{R}^n$ (representada esquemáticamente) desde dos puntos de observación distintos, exactamente como hicimos con la correlación en la lección 5. La primera vista permite apreciar con claridad el triángulo rectángulo de la regresión (hipotenusa $\boldsymbol{y}$ y los dos catetos $\widehat{\boldsymbol{y}}$ y $\widehat{\boldsymbol{e}}$); la segunda, rotada, ayuda a visualizar mejor el ángulo θ entre $\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$, marcado con el arco y protagonista de la próxima transparencia.

Nótese, para quien repase las figuras de la lección 6, que aquí *no* se muestra explícitamente la componente $\hat{\beta}_2\boldsymbol{x}$ dentro del plano: solo destaca en él la recta $\mathcal{L}(\mathbf{1})$ y las proyecciones sobre ella. Esto es deliberado: hace que la figura —y el argumento que construimos hoy— sirva igualmente como esquema de la regresión con cualquier número k de regresores, sustituyendo el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ por un subespacio de dimensión k. Volveremos sobre esto en la transparencia "Recapitulación y guiño a la lección siguiente".

Para profundizar:

- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: descomposición ortogonal y teorema de Pitágoras en $\mathbb{R}^n$.

3 Pitágoras: la varianza se descompone

Aplicando el teorema de Pitágoras (lección 3) sobre $\boldsymbol{y} - \overline{\boldsymbol{y}} = (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}}$, tenemos:

$$\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 \;=\; \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2 \;+\; \|\widehat{\boldsymbol{e}}\|_s^2.$$

Es decir: $\boxed{\sigma_{\boldsymbol{y}}^2 \;=\; \sigma_{\widehat{\boldsymbol{y}}}^2 \;+\; \sigma_{\widehat{\boldsymbol{e}}}^2}$,

donde hemos usado que $\overline{\boldsymbol{y}} = \mu_{\boldsymbol{y}}\mathbf{1} = \mu_{\widehat{\boldsymbol{y}}}\mathbf{1} = \overline{\widehat{\boldsymbol{y}}}$ (pues $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$).

Y multiplicando por n (es decir, volviendo a la norma euclidea usual en $\mathbb{R}^n$) tenemos:

$$\underbrace{\textstyle\sum_i (y_i - \mu_{\boldsymbol{y}})^2}_{\text{STC}} \;=\; \underbrace{\textstyle\sum_i (\hat{y}_i - \mu_{\boldsymbol{y}})^2}_{\text{SEC}} \;+\; \underbrace{\textstyle\sum_i \hat{e}_i^2}_{\text{SRC}};$$

que es una expresión habitual en los manuales de econometría (o programas econométricos como Gretl) donde: STC = Suma Total de Cuadrados, SEC = Suma Explicada de Cuadrados y SRC = Suma Residual de Cuadrados.

Pitágoras: la varianza se descompone

Con el triángulo rectángulo ya establecido en la transparencia anterior, aplicar el teorema de Pitágoras[1] es inmediato:

$$\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 = \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2 + \|\widehat{\boldsymbol{e}}\|_s^2.$$

Reconociendo que $\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2 = \sigma_{\boldsymbol{y}}^2$ (definición de varianza, lección 5) y que $\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2 = \sigma_{\widehat{\boldsymbol{y}}}^2$ (la varianza del propio vector ajustado puesto que $\overline{\boldsymbol{y}} = \mu_{\boldsymbol{y}}\mathbf{1} = \mu_{\widehat{\boldsymbol{y}}}\mathbf{1} = \overline{\widehat{\boldsymbol{y}}}$), llegamos a

$$\sigma_{\boldsymbol{y}}^2 = \sigma_{\widehat{\boldsymbol{y}}}^2 + \sigma_{\widehat{\boldsymbol{e}}}^2;$$

donde, como $\widehat{\boldsymbol{e}}$ tiene media cero, $\mu_{\widehat{\boldsymbol{e}}} = 0$ (primera condición de ortogonalidad de la lección 7), su varianza $\sigma_{\widehat{\boldsymbol{e}}}^2$ coincide con $\|\widehat{\boldsymbol{e}}\|_s^2$, puesto que $\overline{\widehat{\boldsymbol{e}}} = \mathbf{0}$.

Esa es la razón por la que los tres vectores involucrados en el triángulo rectángulo: $(\boldsymbol{y} - \overline{\boldsymbol{y}})$, $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$ y $\widehat{\boldsymbol{e}}$, aparecen, en la figura de la transparencia anterior, en el plano $\mathcal{L}(\mathbf{1})^\perp$ perpendicular a la recta $\mathcal{L}(\mathbf{1})$ (los tres pertenecen al plano de aquellos vectores que tienen media cero).

Sobre el cambio de lenguaje a STC=SEC+SRC en los manuales de econometría y programas econométricos como Gretl. Aquí hemos usado sistemáticamente el producto escalar de la estadística $\langle \cdot, \cdot \rangle_s$ (con su factor $1/n$) a lo largo de las lecciones 4, 5, 6 y 7. La medición de los vectores en desviaciones $(\boldsymbol{y} - \overline{\boldsymbol{y}})$, $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$ y $\widehat{\boldsymbol{e}}$ usando el producto escalar de la estadística, arroja la identidad $\sigma_{\boldsymbol{y}}^2 = \sigma_{\widehat{\boldsymbol{y}}}^2 + \sigma_{\widehat{\boldsymbol{e}}}^2$. Sin embargo, en la inmensa mayoría de los libros de texto —y en la salida que ofrece Gretl— esta identidad se presenta multiplicada por n (es decir, omitiendo el factor $1/n$):

$$\text{STC} = \text{SEC} + \text{SRC};$$

con $\text{STC} = \sum_i (y_i - \mu_{\boldsymbol{y}})^2$, $\text{SEC} = \sum_i (\hat{y}_i - \mu_{\boldsymbol{y}})^2$ y $\text{SRC} = \sum_i \hat{e}_i^2$.

Es decir, en la literatura es tradicional expresar la identidad pitagórica de más arriba midiendo el cuadrado de las longitudes de los vectores $(\boldsymbol{y} - \overline{\boldsymbol{y}})$, $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$ y $\widehat{\boldsymbol{e}}$ con el producto escalar usual en el espacio euclídeo: $\langle \cdot, \cdot \rangle_e$.

[1] demostrado en la lección 3 para cualquier producto escalar con simetría, linealidad y positividad —propiedades que hereda $\langle \cdot, \cdot \rangle_s$, como se explicitó en la lección 4.

Nota: tenga en cuenta que en inglés la identidad se escribe habitualmente SST = SSE + SSR (Sum of Squares Total, Explained, Residual). Aquí conviene una advertencia, porque es una fuente de confusión frecuente: en esta convención anglosajona, SSE denota la suma *explicada*, no la suma de errores al cuadrado (que sería lo que la sigla "E" sugiere a primera vista). En este curso usaremos las siglas españolas STC/SEC/SRC.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, ecuación (2.35) y comentario adyacente sobre la nomenclatura SST/SSE/SSR.

4 ¿Qué ajuste es mejor: el de arriba o el de abajo?

Mismo $\widehat{\boldsymbol{y}}$ y mismo $\overline{\boldsymbol{y}}$ arriba y abajo; solo cambia $\widehat{\boldsymbol{e}}$. Para comparar $\boldsymbol{y}$ con $\widehat{\boldsymbol{y}}$ basta mirar lo que *no* comparten.

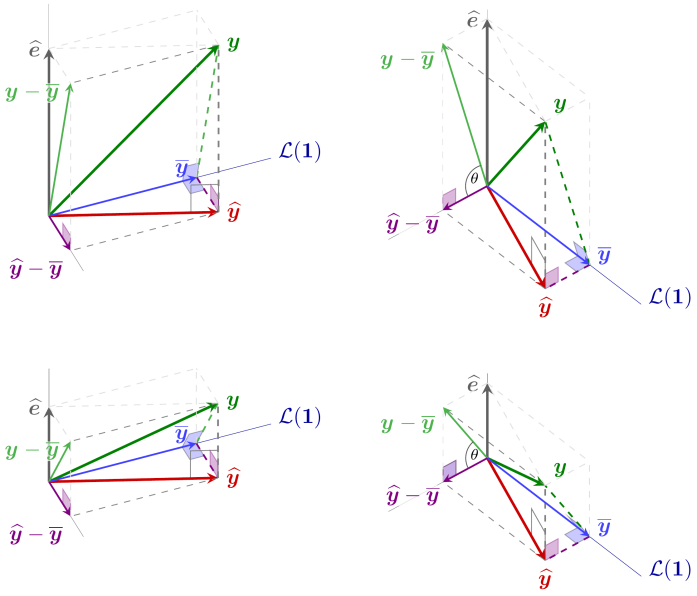

Figura 2: ¿Qué ajuste es mejor, el de la fila de arriba o el de la fila de abajo? Cada fila de gráficos muestra un mismo ajuste desde dos ángulos distintos. Además, los ajustes en ambas filas comparten el mismo vector de medias $\overline{\boldsymbol{y}}$ (azul), el mismo vector de ajuste $\widehat{\boldsymbol{y}}$ (rojo) y, por tanto, el mismo vector en desviaciones del ajuste $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ (violeta). Lo único que cambia es el tamaño del vector de residuos $\widehat{\boldsymbol{e}}$ (gris) —mayor arriba, menor abajo— y, en consecuencia, el propio $\boldsymbol{y}$ (verde oscuro) y su vector en desviaciones $\boldsymbol{y} - \overline{\boldsymbol{y}}$ (verde claro) son distintos arriba y abajo.

¿Qué ajuste es mejor: el de arriba o el de abajo?

Antes de dar ninguna definición formal, merece la pena razonar sobre la figura, porque el razonamiento mismo va a sugerir, de manera casi inevitable, cómo debe definirse la medida de

bondad de ajuste que buscamos.

El ajuste $\widehat{\boldsymbol{y}}$ es mejor cuanto más se parezca a $\boldsymbol{y}$ —dicho con precisión (lección 3), cuanto menor sea la distancia $\|\boldsymbol{y}-\widehat{\boldsymbol{y}}\|_s$—. Pero comparar $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$ directamente esconde un hecho que ya conocíamos desde la lección 7: ambos vectores comparten la misma componente sobre $\mathcal{L}(\mathbf{1})$, el vector de medias $\overline{\boldsymbol{y}}$ (pues $\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}$). Esta es la razón por la que en la figura de arriba y en la de abajo el vector azul es idéntico: sea cual sea el tamaño del residuo, esa parte nunca cambia.

Como esa componente es siempre común, no aporta ninguna información sobre la calidad del ajuste: da igual cuál sea $\boldsymbol{y}$, el vector de medias de $\boldsymbol{y}$ y el vector de medias $\widehat{\boldsymbol{y}}$ son inevitablemente el mismo vector ($\overline{\boldsymbol{y}} = \overline{\widehat{\boldsymbol{y}}}$). Por tanto, para comparar el ajuste basta con mirar lo que *no* es común: los dos vectores en desviaciones, $\boldsymbol{y}-\overline{\boldsymbol{y}}$ (verde claro) y $\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}$ (violeta).

Ya sabemos que estos dos vectores forman parte de un triángulo rectángulo: $\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}$ es un cateto, y $\boldsymbol{y}-\overline{\boldsymbol{y}}$ es la hipotenusa (el otro cateto es $\widehat{\boldsymbol{e}}$). Por tanto, $\|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s \leq \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s$ (pues un cateto nunca es más largo que la hipotenusa), con igualdad exactamente cuando $\widehat{\boldsymbol{e}} = \mathbf{0}$ (ajuste perfecto cuando los dos vectores en desviaciones coinciden).

Esto sugiere de manera natural cómo medir la bondad del ajuste: comparando las longitudes de esos dos vectores en desviaciones, por ejemplo, mirando el ratio

$$\frac{\|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s^2}{\|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s^2}$$

(elevado al cuadrado, para trabajar con varianzas en lugar de con desviaciones típicas). Si $\widehat{\boldsymbol{e}}$ es pequeño (figura de abajo), las dos longitudes son casi iguales, y el ratio está cerca de 1: buen ajuste. Si $\widehat{\boldsymbol{e}}$ es grande (figura de arriba), $\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}$ es mucho más corto que $\boldsymbol{y}-\overline{\boldsymbol{y}}$, y el ratio se hace mucho más pequeño que 1: mal ajuste. En nuestra figura, por tanto, *el ajuste de abajo es mejor que el de arriba* —y el ratio que acabamos de construir lo confirmará numéricamente en la próxima transparencia, donde por fin lo llamamos por su nombre: el coeficiente de determinación R^2.

5 El coeficiente de determinación R^2

El *coeficiente de determinación* es el cociente entre las normas al cuadrado del ajuste y de los datos, ambos en desviaciones:

$$R^2 \;=\; \frac{\|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s^2}{\|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s^2} \;=\; \frac{\sigma^2_{\widehat{\boldsymbol{y}}}}{\sigma^2_{\boldsymbol{y}}}.$$

Multiplicando por n numerador y denominador: $R^2 = \mathrm{SEC}/\mathrm{STC} = 1 - \mathrm{SRC}/\mathrm{STC}$.

Del triángulo rectángulo, como $\langle \boldsymbol{y}-\overline{\boldsymbol{y}}, \widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\rangle_s = \|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s^2$ (pues $\widehat{\boldsymbol{e}} \perp \mathcal{L}(\mathbf{1}, \boldsymbol{x})$): $\cos\theta = \|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s / \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s = \sigma_{\widehat{\boldsymbol{y}}}/\sigma_{\boldsymbol{y}} = \rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}}$.

$$\boxed{R^2 \;=\; \cos^2\theta \;=\; \rho^2_{\widehat{\boldsymbol{y}}\boldsymbol{y}}} \quad \text{donde } \theta = \measuredangle\Big((\boldsymbol{y}-\overline{\boldsymbol{y}}),\, (\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}})\Big);$$

como todo coseno al cuadrado, $0 \leq R^2 \leq 1$.

El coeficiente de determinación: $R^2 = \frac{\|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_s^2}{\|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s^2}$

La transparencia anterior nos dejó el camino casi completo: la bondad del ajuste se puede medir comparando las longitudes de los vectores en desviaciones $\boldsymbol{y}-\overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}$. Formalizamos esa intuición como definición:

Definición. El coeficiente de determinación es

$$R^2 = \frac{\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2}{\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s^2} = \frac{\sigma^2_{\widehat{\boldsymbol{y}}}}{\sigma^2_{\boldsymbol{y}}}.$$

Multiplicando numerador y denominador por n se obtiene la forma habitual en los manuales:

$$R^2 = \frac{\mathrm{SEC}}{\mathrm{STC}} = 1 - \frac{\mathrm{SRC}}{\mathrm{STC}};$$

ya que, como $\mathrm{STC} = \mathrm{SEC} + \mathrm{SRC}$, podemos sustituir $\mathrm{SEC} = \mathrm{STC} - \mathrm{SRC}$.[2]

Esta definición como ratio de normas admite una segunda lectura, tan reveladora como la primera: es, literalmente, el coseno al cuadrado de un ángulo. Recordemos la definición de coseno entre dos vectores (lección 3) aplicada aquí a los vectores en desviaciones de $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$ (como se hizo con $\boldsymbol{x}$ e $\boldsymbol{y}$ en la lección 5): el ángulo θ entre $\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ cumple

$$\cos\theta \;=\; \frac{\langle \boldsymbol{y} - \overline{\boldsymbol{y}},\, \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \rangle_s}{\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s \, \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s} \;=\; \rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}}; \tag{1}$$

donde hemos usado la definición de correlación de la lección 5.

Para simplificar el numerador de (1), sustituimos $\boldsymbol{y}-\overline{\boldsymbol{y}} = (\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}})+\widehat{\boldsymbol{e}}$ (transparencia "El triángulo rectángulo de la bondad de ajuste"):

$$\langle \boldsymbol{y} - \overline{\boldsymbol{y}},\, \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \rangle_s = \left\langle \left((\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}} \right),\, \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \right\rangle_s = \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2 + \underbrace{\langle \widehat{\boldsymbol{e}}, \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \rangle_s}_{=0 \;(\widehat{\boldsymbol{e}} \perp \mathcal{L}(\mathbf{1}, \boldsymbol{x}))} = \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2;$$

donde hemos usado que $\widehat{\boldsymbol{e}}$ es ortogonal al plano que contiene a los vectores $\widehat{\boldsymbol{y}}$ y $\overline{\boldsymbol{y}}$.

Sustituyendo:

$$\cos\theta = \frac{\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s^2}{\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s \, \|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s} = \frac{\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|_s}{\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|_s} = \frac{\sigma_{\widehat{\boldsymbol{y}}}}{\sigma_{\boldsymbol{y}}}.$$

Elevando al cuadrado, y comparando con la definición del principio de esta misma transparencia:

$$\cos^2\theta = \frac{\sigma^2_{\widehat{\boldsymbol{y}}}}{\sigma^2_{\boldsymbol{y}}} = R^2,$$

donde, además, $\cos\theta = \rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}}$.

[2] Fíjese que $\frac{\mathrm{SEC}}{\mathrm{STC}} = \frac{\|\widehat{\boldsymbol{y}}-\overline{\boldsymbol{y}}\|_e^2}{\|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_e^2}$ es el ratio de normas al cuadrado de los mismos vectores en desviaciones, pero medidos usando el producto escalar habitual $\langle\cdot,\cdot\rangle_e$ del espacio euclídeo $\mathbb{R}^n$

Las dos lecturas son, por tanto, la misma cosa vista de dos maneras. R^2 es un ratio de normas al cuadrado (la forma que motivamos con la figura de la transparencia anterior) y es, a la vez, un coseno al cuadrado (la forma habitual en los manuales, que conecta con el cuadrado de la correlación entre el vector de datos $\boldsymbol{y}$ y su ajuste $\widehat{\boldsymbol{y}}$). El ángulo θ es el marcado con el arco en la figura de la transparencia "El triángulo rectángulo de la bondad de ajuste": el ángulo entre el vector en desviaciones de los datos y el vector en desviaciones de su ajuste: $\theta = \measuredangle\Big((\boldsymbol{y} - \overline{\boldsymbol{y}}),\, (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})\Big)$.

El hecho de que $0 \leq R^2 \leq 1$ tiene *dos* justificaciones independientes, que conviene señalar porque es un ejemplo bonito de cómo dos caminos distintos llevan a la misma cota. Por un lado, ya lo sabíamos por Pitágoras (transparencia "Pitágoras: la varianza se descompone"): como SEC+SRC = STC y SRC ≥ 0, forzosamente SEC $\leq$ STC, luego $R^2 \leq 1$ (y $R^2 \geq 0$ porque SEC y SRC son el cuadrado de dos longitudes o normas). Por otro, ahora lo sabemos también porque ningún coseno se sale de $[-1, 1]$, así que ningún coseno al cuadrado se sale de $[0, 1]$ —sin necesidad de invocar Pitágoras.

6 Interpretando R^2: casos extremos

Con $\boldsymbol{y} = \widehat{\boldsymbol{y}}$:

- $\theta = 0°$ ($\widehat{\boldsymbol{e}} = \boldsymbol{0}$): el ajuste *coincide* con los datos. $R^2 = 1$.

Con $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})$ fijo y no nulo, R^2 nunca *llega* a ser 0: solo se le acerca.

- $\theta \to 90°$: si el residuo crece *sin límite*; $R^2 \to 0$ (solo 0 como límite).

¿Y puede R^2 ser exactamente 0?

- Sí — pero en otra situación: cuando el *propio ajuste* coincide con el vector de medias, $\widehat{\boldsymbol{y}} = \overline{\boldsymbol{y}}$ (cateto nulo, sin residuo infinito). Es decir, cuando $\rho_{\boldsymbol{xy}} = 0$.

Interpretando R^2: casos extremos

Las figuras anteriores permiten comprender, sin necesidad de ninguna fórmula adicional, cuando se dan los casos extremos.

Si $\widehat{\boldsymbol{e}} = \boldsymbol{0}$, el triángulo rectángulo degenera: la hipotenusa $\boldsymbol{y} - \overline{\boldsymbol{y}}$ coincide con el cateto $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$, el ángulo θ es $0°$, y $R^2 = \cos^2 0° = 1^2$: el ajuste reproduce los datos exactamente. Si, en cambio, dejamos crecer $\widehat{\boldsymbol{e}}$ sin límite (manteniendo fijo el cateto $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$, como en la figura), la hipotenusa se alarga cada vez más en la dirección de $\mathcal{L}(\boldsymbol{1}, \boldsymbol{x})^{\perp}$, el ángulo θ se abre progresivamente hacia $90°$, y $R^2 = \cos^2\theta$ se acerca a 0. Nótese que decir "θ se abre hacia $90°$" es, aquí, lo mismo que decir "el residuo crece sin límite".

Conviene precisar un matiz que la propia figura deja entrever. Mientras el cateto $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ permanezca fijo y sea *distinto de* $\boldsymbol{0}$ (como en las dos escenas de la figura), $R^2 = \sigma^2_{\widehat{\boldsymbol{y}}}/(\sigma^2_{\widehat{\boldsymbol{y}}} + \sigma^2_{\widehat{\boldsymbol{e}}})$ es siempre estrictamente positivo, por grande que sea el residuo: R^2 se *acerca* a 0 tanto como queramos, pero solo lo *alcanza* en el límite de un residuo de norma infinita, que no es una situación real.

Esto no significa que R^2 no pueda ser *exactamente* 0 en la práctica: sí puede, pero en una situación distinta de la que muestra esta figura. Ocurre cuando el propio ajuste coincide con el

vector de medias, $\widehat{\boldsymbol{y}} = \overline{\boldsymbol{y}}$ —es decir, cuando el cateto $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ es $\mathbf{0}$, no cuando el residuo se dispara—. En la regresión simple, esto sucede exactamente cuando $\hat{\beta}_2 = 0$ (covarianza muestral nula entre $\boldsymbol{x}$ e $\boldsymbol{y}$, lección 7): el mejor ajuste posible resulta ser, sencillamente, "predecir siempre la media", sin que $\boldsymbol{x}$ aporte nada.[3] Es importante no confundir esta situación —perfectamente posible con datos $\boldsymbol{y}$ ordinarios, no constantes— con el caso degenerado en que $\boldsymbol{y}$ mismo fuera constante: en ese caso $\sigma_{\boldsymbol{y}}^2 = 0$ y R^2 ni siquiera estaría definido (0/0), igual que ocurría con la correlación en la lección 5.

7 Ejemplo numérico: STC, SEC, SRC

Retomamos el ejemplo de la lección 7: $\boldsymbol{x} = (1, 2, 3, 4, 5)$, $\boldsymbol{y} = (6, 6, 11, 16, 16)$, con $\hat{\beta}_1 = 2$, $\hat{\beta}_2 = 3$.

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

$\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}} = 11$. Entonces:

$$\text{STC} = \sum (y_i - 11)^2 = 100, \qquad \text{SEC} = \sum (\hat{y}_i - 11)^2 = 90, \qquad \text{SRC} = \sum \hat{e}_i^2 = 10.$$

Comprobación: $90 + 10 = 100$. $R^2 = 90/100 = 0{,}9$.

Con $\sigma_{\boldsymbol{x}}^2 = 2$, $\sigma_{\boldsymbol{y}}^2 = 20$, $\sigma_{\boldsymbol{xy}} = 6$: $\rho_{\boldsymbol{xy}}^2 = 6^2/(2 \cdot 20) = 36/40 = 0{,}9$. ¡Coincide!

Ejemplo numérico: STC, SEC, SRC

Verifiquemos con números todo lo dicho, retomando exactamente el ejemplo de la lección 7 (el mismo $n = 5$, con $\boldsymbol{x} = (1, 2, 3, 4, 5)$, $\boldsymbol{u} = (1, -2, 0, 2, -1)$ y $\boldsymbol{y} = 2 + 3\boldsymbol{x} + \boldsymbol{u} = (6, 6, 11, 16, 16)$), donde ya obtuvimos $\hat{\beta}_1 = 2$, $\hat{\beta}_2 = 3$ y comprobamos que, en ese ejemplo con truco, $\widehat{\boldsymbol{e}} = \boldsymbol{u}$.

Los valores ajustados son $\hat{y}_i = 2 + 3x_i$: $\widehat{\boldsymbol{y}} = (5, 8, 11, 14, 17)$. Los residuos, $\hat{e}_i = y_i - \hat{y}_i$: $\widehat{\boldsymbol{e}} = (1, -2, 0, 2, -1)$ (efectivamente, coincide con $\boldsymbol{u}$, como ya sabíamos).

Media de $\boldsymbol{y}$ *y de* $\widehat{\boldsymbol{y}}$: $\mu_{\boldsymbol{y}} = \frac{6+6+11+16+16}{5} = 11$; $\mu_{\widehat{\boldsymbol{y}}} = \frac{5+8+11+14+17}{5} = \frac{55}{5} = 11$. Coinciden, como debía ser.

STC: $\sum_i (y_i - 11)^2 = (-5)^2 + (-5)^2 + 0^2 + 5^2 + 5^2 = 25 + 25 + 0 + 25 + 25 = 100$.

SEC: $\sum_i (\hat{y}_i - 11)^2 = (-6)^2 + (-3)^2 + 0^2 + 3^2 + 6^2 = 36 + 9 + 0 + 9 + 36 = 90$.

SRC: $\sum_i \hat{e}_i^2 = 1^2 + (-2)^2 + 0^2 + 2^2 + (-1)^2 = 1 + 4 + 0 + 4 + 1 = 10$.

Comprobación de Pitágoras: $\text{SEC} + \text{SRC} = 90 + 10 = 100 = \text{STC}$.

[3] Nótese una sutileza de dominio, coherente con la advertencia de la lección 5 sobre la correlación de un vector constante: en este caso límite $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} = \mathbf{0}$, así que $\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}}$ —al ser el coseno de un ángulo con un vector nulo— deja de estar definida. El propio R^2, en cambio, sigue perfectamente definido: es simplemente el cociente $0/\sigma_{\boldsymbol{y}}^2 = 0$. La identidad $R^2 = \rho^2_{\widehat{\boldsymbol{y}}\boldsymbol{y}}$ debe leerse, por tanto, con esta salvedad: el lado izquierdo siempre está definido; el derecho, no siempre.

$R^2 = \text{SEC}/\text{STC} = 90/100 = 0{,}9$: el ajuste explica el 90 % de la variación total de $\boldsymbol{y}$.

Verifiquemos también la vía de la correlación (transparencia anterior). De la lección 7 ya teníamos $\sigma_{\boldsymbol{x}}^2 = 2$ y $\sigma_{\boldsymbol{xy}} = 6$. Y $\sigma_{\boldsymbol{y}}^2 = \text{STC}/n = 100/5 = 20$. Entonces

$$\rho_{\boldsymbol{xy}}^2 = \frac{\sigma_{\boldsymbol{xy}}^2}{\sigma_{\boldsymbol{x}}^2\,\sigma_{\boldsymbol{y}}^2} = \frac{6^2}{2 \cdot 20} = \frac{36}{40} = 0{,}9.$$

Coincide con R^2: una comprobación numérica anticipada del resultado general $R^2 = \rho_{\boldsymbol{xy}}^2$, que demostraremos formalmente más adelante, en la transparencia "En el caso de la regresión simple: $R^2 = \rho_{\boldsymbol{xy}}^2$".

8 Una advertencia: el nombre *suma explicada* es engañoso

STC = SEC + SRC es una identidad *puramente geométrica* sobre vectores de $\mathbb{R}^n$: se cumple siempre, contengan esos vectores lo que contengan.

Primer peligro: un R^2 alto NO requiere ninguna relación real entre $\boldsymbol{x}$ e $\boldsymbol{y}$.

Ejemplo célebre (Vigen, *Spurious Correlations*): consumo de queso per cápita en EE.UU. y personas que mueren enredadas en sus sábanas, $\rho \approx 0{,}947$. Nadie propone que el queso mate sofocando a sus víctimas entre las sábanas.

Segundo peligro: aun cuando SÍ existe una relación real, R^2 no dice nada sobre su *dirección*.

Ampliar una vivienda aumenta su precio (relación razonable). Pero si regresamos *superficie sobre precio* (invirtiendo los papeles), R^2 *no cambia* — y de ahí no se sigue que cuando el precio sube la casa aumenta su tamaño.

Un R^2 alto no certifica relación alguna; y cuando la hay, no indica su sentido.

Solo cuando el modelo sea sensato un R^2 alto será una buena señal.

Una advertencia: el nombre *suma explicada* es engañoso

Antes de seguir, conviene detenerse en una advertencia que atañe al propio *nombre* que hemos venido usando: "Suma *Explicada* de Cuadrados". La identidad STC = SEC + SRC —o, equivalentemente, $\sigma_{\boldsymbol{y}}^2 = \sigma_{\widehat{\boldsymbol{y}}}^2 + \sigma_{\widehat{\boldsymbol{e}}}^2$— es una consecuencia puramente geométrica de la descomposición ortogonal $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}$. Se cumple independientemente del contenido de los vectores $\boldsymbol{x}$ e $\boldsymbol{y}$: la geometría de $\mathbb{R}^n$ no sabe nada de economía, de causalidad, ni de sentido común. El nombre "explicada" sugiere que hay algo sustantivo detrás; la propia matemática ni lo sugiere ni lo garantiza en modo alguno.

Primer peligro: no hace falta ninguna relación real para obtener un R^2 alto. El ejemplo más elocuente es el de las llamadas *correlaciones espurias*. Tyler Vigen, en su proyecto *Spurious Correlations* (tylervigen.com), recopiló decenas de pares de series con correlaciones altísimas y sin ningún vínculo causal plausible. Un ejemplo: el consumo de queso per cápita en Estados Unidos y el número de personas que mueren cada año enredadas en sus sábanas, con $\rho \approx 0{,}947$ durante el periodo 2000-2009 (un $R^2 \approx 0{,}897$ si regresáramos una serie sobre la otra). Nadie sostendría que el consumo de queso provoca —ni evita— morir enredado en las sábanas: ambas series, sencillamente, crecieron en paralelo por razones completamente ajenas entre sí. La identidad STC = SEC + SRC se

cumple, con toda su fuerza matemática, igual para estos dos vectores sin sentido que para cualquier par de vectores con un significado económico razonable: la matemática, por sí sola, no distingue un caso del otro.

Segundo peligro, distinto del anterior: incluso cuando la relación SÍ es real, el R^2 no dice nada sobre su dirección. Retomemos el ejemplo del precio y la superficie de las viviendas. Aquí sí existe una relación económicamente razonable: ampliar la superficie de una vivienda tiende a aumentar su precio. Si regresamos precio sobre superficie, obtenemos un R^2 que, en efecto, sugiere que el modelo parece "explicar" el precio a partir del tamaño. Pero si invertimos los papeles —regresando *superficie* sobre /precio/— obtenemos el mismo R^2. La razón la veremos en la próxima transparencia: en la regresión lineal simple $R^2 = \rho^2_{\boldsymbol{xy}}$, y la correlación no distingue cuál de las dos variables hace de x y cuál de y. Y sin embargo, decir que "el precio explica la superficie" —en el sentido de que cuando los precios de las viviendas suben hace que sus superficies crezcan— no tiene ningún sentido. La regresión, y el R^2 que de ella se deriva, son completamente indiferentes a cuál de las dos variables juega el papel de causa y cuál el de efecto: esa es una pregunta que solo puede responder una teoría económica previa, nunca la mera geometría de los datos.

La lección que hay que extraer de ambos peligros es la misma, aunque se manifieste de dos maneras distintas: *un R^2 alto solo puede ser tomado como "una buena señal" cuando el modelo es sensato y está bien fundamentado desde un punto de vista teórico.* Un R^2 alto no sirve, ni para afirmar que existe una relación real (primer peligro), ni para afirmar en qué sentido va esa relación si de verdad existe (segundo peligro). Tampoco un R^2 bajo implica que el modelo deba ser descartado. Retomaremos esta discusión, con más herramientas, en la lección 9 ("El puente").

> *Advertencia para más adelante.* R^2 tiene una propiedad puramente mecánica que conviene no olvidar: *nunca empeora al añadir regresores*, aunque esos regresores aporten muy poca información real. Precisamente por eso, más adelante necesitaremos una versión corregida de esta medida: el R^2 *ajustado*.

Conviene subrayar una diferencia con el ejemplo de correlación tratado en la lección 5. Allí, $\rho_{\boldsymbol{xy}} = \cos\theta'$ era el coseno del ángulo entre los vectores en desviaciones, $\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\boldsymbol{x} - \overline{\boldsymbol{x}}$, correspondientes a los *datos* y el *regresor*. Aquí, $R^2 = \rho^2_{\boldsymbol{y}\widehat{\boldsymbol{y}}} = \cos^2\theta$ es el cuadrado del coseno del ángulo *entre los datos y su ajuste* ($\boldsymbol{y} - \overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$), *ambos en desviaciones* (si no, no sería una correlación). Son, en principio, dos ejemplos de ángulos distintos, definidos entre parejas de vectores distintas. Sin embargo, en la próxima transparencia veremos que, en el caso particular de la regresión lineal simple (un único regresor $\boldsymbol{x}$ además de la constante), ambos ángulos coinciden.

Para profundizar:

- Vigen, T. *Spurious Correlations*, tylervigen.com: catálogo de correlaciones estadísticamente altas y causalmente vacías.
- Wooldridge, J. M. (2020), cap. 2, sección 2.4: advertencias sobre la interpretación causal de R^2.

9 En el caso de la regresión simple: $R^2 = \rho^2_{\boldsymbol{xy}}$

El ajuste $\widehat{\boldsymbol{y}} = \hat{\beta}_2\boldsymbol{x} + \hat{\beta}_1\mathbf{1}$ es una *traslación y reescalado* de $\boldsymbol{x}$. Aplicando las propiedades de traslación y escala de la correlación:

$$\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}} \;=\; \frac{\hat{\beta}_2}{|\hat{\beta}_2|}\,\rho_{\boldsymbol{xy}}.$$

Como $\hat{\beta}_2$ y $\rho_{\boldsymbol{xy}}$ comparten siempre signo (lección 7): $\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}} = |\rho_{\boldsymbol{xy}}| \geq 0$.

$$\boxed{R^2 \;=\; \rho^2_{\widehat{\boldsymbol{y}}\boldsymbol{y}} \;=\; \rho^2_{\boldsymbol{xy}}.}$$

No es casualidad: en $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ solo hay *una* dirección perpendicular $\mathbf{1}$. Todo vector no constante del plano —incluido $\widehat{\boldsymbol{y}}$— forma con $\boldsymbol{y}$ el mismo ángulo, salvo quizá el signo.

En el caso de la regresión simple: $R^2 = \rho^2_{\boldsymbol{xy}}$

Esta transparencia cierra la identidad guardada de la lección 7, pero conviene detenerse en *por qué* el resultado es cierto, porque el argumento explica también por qué es exclusivo de la regresión simple.

Empecemos por un hecho general sobre el plano $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, que no depende en absoluto de cuál sea el ajuste MCO. Tomemos *cualquier* vector del plano, es decir, cualquier $\boldsymbol{w} = c_1\mathbf{1} + c_2\boldsymbol{x}$. Su media es $\mu_{\boldsymbol{w}} = c_1 + c_2\mu_{\boldsymbol{x}}$, así que su vector en desviaciones es

$$\boldsymbol{w} - \overline{\boldsymbol{w}} = (c_1\mathbf{1} + c_2\boldsymbol{x}) - (c_1 + c_2\mu_{\boldsymbol{x}})\mathbf{1} = c_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}).$$

Es decir: *el vector en desviaciones de* $\boldsymbol{w}$ *es algún múltiplo de* $\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1}$. Dicho de otro modo, todos los vectores de $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$, una vez centrados, caen en la misma recta $\mathcal{L}(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$ —una única dirección dentro de $\mathcal{L}(\mathbf{1})^{\perp}$—. Esto no es más que la identidad guardada de la lección 7, $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} = \hat{\beta}_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$, generalizada de $\widehat{\boldsymbol{y}}$ a *cualquier* vector del plano: $\widehat{\boldsymbol{y}}$ no tiene nada de especial en este sentido; es, sencillamente, uno más de esos vectores $\boldsymbol{w}$ (con $c_1 = \hat{\beta}_1$, $c_2 = \hat{\beta}_2$), y como tal hereda la propiedad.

Esta observación tiene una consecuencia inmediata sobre los ángulos. Si dos vectores en desviaciones son proporcionales y *no nulos*, $(\boldsymbol{w} - \overline{\boldsymbol{w}}) = c_2(\boldsymbol{x} - \mu_{\boldsymbol{x}}\mathbf{1})$ (con $c_2 \neq 0$), entonces el ángulo que un tercer vector $\boldsymbol{y}$ *no nulo* forma con $(\boldsymbol{w} - \overline{\boldsymbol{w}})$ coincide, salvo un posible cambio de 180°, con el ángulo formado entre $\boldsymbol{y}$ y $(\boldsymbol{x} - \overline{\boldsymbol{x}})$. Y como el coseno de θ y el de $180° - \theta$ solo difieren en el signo, las *correlaciones* $\rho_{\boldsymbol{xy}}$ y $\rho_{\boldsymbol{wy}}$ coinciden en valor absoluto, sea cual sea el vector $\boldsymbol{w}$ elegido (con tal de que $c_2 \neq 0$).

Podemos evitar repetir el argumento desde cero: basta combinar las dos propiedades demostradas en la lección 5 bajo el nombre de *propiedades de traslación y escala* de la correlación. Allí vimos que, si $\boldsymbol{x}' = \lambda\boldsymbol{x} + a\mathbf{1}$ (una traslación seguida de un reescalado), entonces

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

Y el ajuste $\widehat{\boldsymbol{y}} = \hat{\beta}_2 \boldsymbol{x} + \hat{\beta}_1 \mathbf{1}$ es, literalmente, una transformación de ese tipo: un reescalado de $\boldsymbol{x}$ por $\hat{\beta}_2$, seguido de una traslación por $\hat{\beta}_1 \mathbf{1}$. Aplicando la fórmula con $\lambda = \hat{\beta}_2$:

$$\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}} = \frac{\hat{\beta}_2}{|\hat{\beta}_2|}\, \rho_{\boldsymbol{x}\boldsymbol{y}}.$$

No hace falta ninguna cuenta nueva: es la propiedad de la lección 5, aplicada al caso concreto en que el "reescalado" es precisamente la pendiente de la regresión.

Solo falta un ingrediente, ya demostrado en la lección 7: el signo de $\hat{\beta}_2$ coincide siempre con el signo de $\rho_{\boldsymbol{x}\boldsymbol{y}}$ (porque $\hat{\beta}_2 = \sigma_{\boldsymbol{x}\boldsymbol{y}}/\sigma^2_{\boldsymbol{x}}$ y $\sigma^2_{\boldsymbol{x}} > 0$). Así que el factor $\hat{\beta}_2/|\hat{\beta}_2|$ es exactamente $\operatorname{sign}(\rho_{\boldsymbol{x}\boldsymbol{y}})$, y

$$\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}} = \operatorname{sign}(\rho_{\boldsymbol{x}\boldsymbol{y}})\rho_{\boldsymbol{x}\boldsymbol{y}} = |\rho_{\boldsymbol{x}\boldsymbol{y}}|.$$

Elevando al cuadrado, y usando $R^2 = \rho^2_{\widehat{\boldsymbol{y}}\boldsymbol{y}}$ (transparencia "El coeficiente de determinación"):

$$R^2 = \rho^2_{\boldsymbol{x}\boldsymbol{y}}.$$

Este resultado explica *por qué* es exclusivo de la regresión simple, sin necesidad de comparar dos cálculos algebraicos: en $\mathcal{L}(\mathbf{1}, \boldsymbol{x})$ hay una sola dirección perpendicular a los vectores constantes, así que *todos* los vectores no constantes del plano —$\boldsymbol{x}$, $2\boldsymbol{x}$, $\boldsymbol{x} + 7\mathbf{1}$, o el propio $\widehat{\boldsymbol{y}}$— una vez centrados, forman con $\boldsymbol{y}$ el mismo ángulo ($\pm 180°$). En la lección 10 (regresión múltiple con dos o más regresores no constantes ($k > 2$)), el subespacio análogo —los vectores no constantes de $\mathcal{L}(\mathbf{1}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$, centrados— tiene dimensión $k-1$, no 1: ya no hay una única dirección perpendicular a los vectores constantes, y $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ deja de estar alineado con ningún regresor individual centrado. Por eso $R^2 = \cos^2\theta$ seguirá siendo válido (no depende de la dimensión del subespacio), pero dejará de coincidir con el cuadrado de ninguna correlación simple $\rho_{\boldsymbol{x}_j\boldsymbol{y}}$.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, ecuación (2.38): $R^2 = \hat{\rho}^2_{xy}$ en el modelo de regresión simple.

10 Regresión lineal simple: STC, SEC y SRC con áreas

Versión interactiva: cuaderno 3 en MyBinder, parte 2 (las tres áreas, con datos editables).

Es la misma identidad STC = SRC + SEC de la transparencia "Pitágoras: la varianza se descompone", ahora vista como suma de áreas, dato a dato.

Regresión lineal simple: STC, SEC y SRC con áreas

Los tres diagramas de esta figura viven en un espacio distinto del de las demás figuras de esta sesión. No es $\mathbb{R}^n$, el espacio abstracto de los n datos donde vive el triángulo de las transparencias "El triángulo rectángulo de la bondad de ajuste" y "Pitágoras: la varianza se descompone"; es el plano (x, y) de los diagramas de dispersión de toda la vida, que ya usamos en el laboratorio que sigue a la lección 5. Es un complemento de la representación abstracta con vectores: mientras que allí una norma al cuadrado es "la longitud de un vector, elevada al cuadrado", aquí cada sumando de STC, SEC o SRC es, literalmente, *el área de un cuadrado* dibujado sobre el propio diagrama de dispersión.

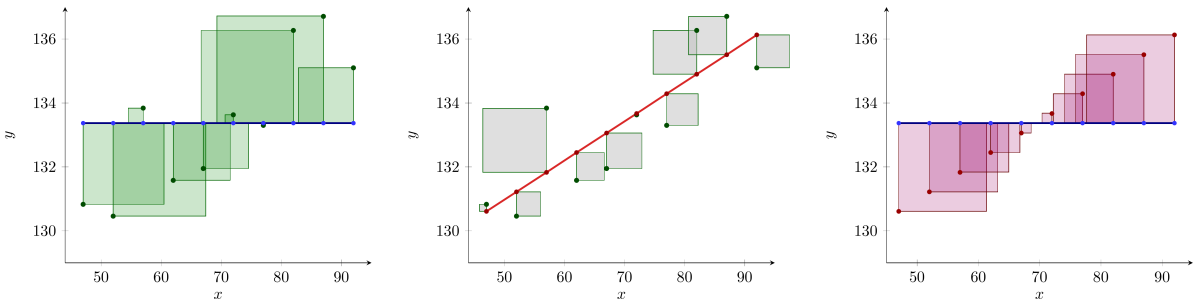

Figura 3: STC, SEC y SRC como áreas de cuadrados sobre el diagrama de dispersión clásico. (Izquierda) STC: cada cuadrado tiene lado $y_i - \mu_{\boldsymbol{y}}$ (distancia del dato a la media). (Centro) SRC: los cuadrados, mucho más pequeños, tienen lado $y_i - \hat{y}_i$ (distancia de cada dato a su ajuste). (Derecha) SEC: los puntos son ahora los ajustes $\hat{y}_i$ (alineados en la recta de regresión); el lado de cada cuadrado es $\hat{y}_i - \mu_{\boldsymbol{y}}$.

En el panel de la izquierda, cada observación (x_i, y_i) aparece en una esquina de un cuadrado, cuyo lado vertical muestra la distancia $|y_i - \mu_{\boldsymbol{y}}|$ entre el dato y la recta horizontal de altura $\mu_{\boldsymbol{y}}$. La suma de las áreas de todos esos cuadrados es exactamente $\text{STC} = \sum_i (y_i - \mu_{\boldsymbol{y}})^2$.

En el panel central, los cuadrados —notablemente más pequeños que en los otros dos paneles, cuando el ajuste es razonablemente bueno— tienen como lado la distancia vertical entre cada dato y su propio ajuste, $|y_i - \hat{y}_i|$, es decir, el residuo $|\hat{e}_i|$. La suma de estas áreas es $\text{SRC} = \sum_i \hat{e}_i^2$.

En el panel de la derecha, los puntos ya no son los datos originales, sino los *valores ajustados* $\hat{y}_i$ —por eso aparecen perfectamente alineados sobre la recta de regresión—; el lado de cada cuadrado es ahora la distancia $|\hat{y}_i - \mu_{\boldsymbol{y}}|$. La suma de áreas de estos cuadrados es $\text{SEC} = \sum_i (\hat{y}_i - \mu_{\boldsymbol{y}})^2$.

En las transparencias "El triángulo rectángulo de la bondad de ajuste" y "Pitágoras: la varianza se descompone" demostramos la identidad $\text{STC} = \text{SEC} + \text{SRC}$ como una única igualdad entre normas de vectores en $\mathbb{R}^n$. Aquí se lee observación a observación: la suma de las áreas de la izquierda es igual a la suma de las áreas del centro más la de la derecha. Es la misma identidad, vista con los ojos del diagrama de dispersión en lugar de los del espacio abstracto de datos. Por eso usamos aquí las siglas STC/SEC/SRC, la versión "sin dividir por n" de la que advertimos en la transparencia "Pitágoras: la varianza se descompone": es la que corresponde a sumar áreas de cuadrados sin más.

Esta segunda representación —cuadrados sobre el diagrama de dispersión frente a vectores abstractos en $\mathbb{R}^n$— es la ilustración clásica con la que muchos manuales de econometría presentan R^2 por primera vez. Aquí llega en último lugar, a propósito: primero conviene entender *por qué* es cierta la identidad (Pitágoras en $\mathbb{R}^n$, transparencias "El triángulo rectángulo de la bondad de ajuste" y "Pitágoras: la varianza se descompone"), y solo después verla como la receta de "sumar áreas de cuadraditos". Además, esta representación gráfica solo es viable y legible en la regresión lineal simple.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2: representación gráfica clásica de la descomposición de va-

rianzas sobre el diagrama de dispersión.

11 Recapitulación y guiño a la lección siguiente

Paso	Resultado
Descomposición ortogonal	$(\boldsymbol{y} - \overline{\boldsymbol{y}}) = (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \widehat{\boldsymbol{e}}$, sumandos ortogonales
Pitágoras	$\sigma^2_{\boldsymbol{y}} = \sigma^2_{\widehat{\boldsymbol{y}}} + \sigma^2_{\widehat{\boldsymbol{e}}}$ (STC=SEC+SRC)
Definición	$R^2 = \sigma^2_{\widehat{\boldsymbol{y}}}/\sigma^2_{\boldsymbol{y}} = \text{SEC/STC}$ $(= \cos^2\theta = \rho^2_{\widehat{\boldsymbol{y}}\boldsymbol{y}})$
Caso simple (un regresor)	$R^2 = \rho^2_{\boldsymbol{xy}}$

Más adelante: el mismo argumento —Pitágoras + las dos ortogonalidades de la lección 7— se generalizará a k regresores: solo cambiará la dimensión del subespacio sobre el que proyectamos. Las primeras figuras de esta sesión, al no mostrar $\hat{\beta}_2\boldsymbol{x}$ explícitamente, valdrán como esquema para el caso general.

Próxima sesión: laboratorio; verificaremos con datos reales que $\sum \hat{e}_i = 0$, $\sum x_i\hat{e}_i = 0$ y que el R^2 calculado a mano coincide con el que reporta Gretl.

Lección 9 — El puente: hasta ahora, pura geometría en $\mathbb{R}^n$, sin ningún supuesto probabilístico. La próxima lección teórica da el salto —como analogía, no como demostración— al espacio de las variables aleatorias: la esperanza condicional $E[Y \mid X]$ jugará el papel de $\widehat{\boldsymbol{y}}$.

Recapitulación y guiño a la lección 9

El recorrido de hoy ha sido, otra vez, muy corto en ingredientes nuevos: la única pieza matemática genuinamente nueva ha sido aplicar el teorema de Pitágoras (lección 3) al triángulo rectángulo formado por $\boldsymbol{y} - \overline{\boldsymbol{y}}$, $\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}$ y $\widehat{\boldsymbol{e}}$ —un triángulo que ya estaba, en germen, en las dos condiciones de ortogonalidad resueltas en la lección 7—. Todo lo demás ha sido, una vez más, *aplicar* herramientas ya conocidas: la definición de varianza (lección 5), la definición de coseno (lección 3), y la fórmula de $\hat{\beta}_2$ (lección 7).

Para cerrar, merece la pena remarcar tres ideas:

1. La descomposición $\sigma^2_{\boldsymbol{y}} = \sigma^2_{\widehat{\boldsymbol{y}}} + \sigma^2_{\widehat{\boldsymbol{e}}}$ (o, sin el factor $1/n$, STC = SEC + SRC) es consecuencia directa e inevitable de las dos ortogonalidades de la lección 7: es lo que se obtiene, sin ningún supuesto adicional, de aplicar Pitágoras a esas ortogonalidades.
2. $R^2 = \cos^2\theta$ mide, literalmente, cuánto se parece —en dirección, no en escala— el vector de datos en desviaciones al vector ajustado en desviaciones. En la regresión simple, ese ángulo coincide con el de la correlación entre $\boldsymbol{x}$ e $\boldsymbol{y}$ (transparencia "En el caso de la regresión simple: $R^2 = \rho^2_{\boldsymbol{xy}}$"), cerrando el círculo abierto en la lección 5; en la regresión múltiple (lección 10) seguirá siendo $\cos^2\theta$, pero ya no será el cuadrado de alguna correlación con un regresor particular.
3. Un R^2 alto no es una condición necesaria para tomarse en serio un modelo, ni un R^2 bajo una condición suficiente para descartarlo. La identidad STC = SEC + SRC es una propiedad geométrica que no distingue entre una relación económicamente sensata y una correlación espuria: eso exige, siempre, una teoría previa ajena a la pura geometría de los datos.

La práctica de laboratorio que sigue a esta lección retoma exactamente los datos con los que se trabajó en la práctica que siguió a la lección 5, y comprobará numéricamente, con un conjunto de datos completo, tanto las dos condiciones de ortogonalidad como la propia descomposición $\text{STC} = \text{SEC} + \text{SRC}$ y el valor de R^2 que reporta Gretl automáticamente al pedir una regresión.

Y la lección 9 ("El puente") dará un paso de naturaleza distinta a los de las lecciones anteriores: hasta ahora hemos trabajado exclusivamente con vectores concretos de $\mathbb{R}^n$, sin ningún supuesto sobre cómo se generaron los datos. La lección 9 presentará, como *analogía* explícita —no como demostración—, el paralelismo entre esta geometría muestral (vectores de datos) y el mundo de las variables aleatorias. El valor esperado jugará el papel de la media; la esperanza $E[\cdot]$, el del vector de medias; la esperanza condicional $E[Y \mid X]$, el de la proyección $\widehat{\boldsymbol{y}}$; y los supuestos del modelo clásico de regresión serán las condiciones que hacen legítimo ese paralelismo. Solo entonces podremos empezar a hablar, con algún rigor, de si cabe una lectura causal de un buen ajuste —la pregunta que hoy hemos dejado, deliberadamente, sin responder.

Para profundizar:

- Wooldridge, J. M. (2020), cap. 2, secciones 2.3-2.4: resumen completo de STC/SEC/SRC, R^2 y sus limitaciones interpretativas.
- Bujosa, M. *Curso de Álgebra Lineal*, cap. 11: teorema de Pitágoras y ángulo entre vectores en $\mathbb{R}^n$.