Matemáticas II

Marcos Bujosa

Universidad Complutense de Madrid

14/01/2023

1 Esquema de la **Lección 19**

Esquema de la Lección 19

- Media
- Desviación típica y varianza
- Mínimos Cuadrados Ordinarios (MCO)

2 Una restrcción en estadística y probabilidad

La norma del vector constante "uno" es 1

Esto no se cumple con el producto punto de $\mathbb{R}^m$ $(m > 1)$

$$\|\mathbf{1}\|^2 = \langle \mathbf{1}, \mathbf{1} \rangle = \mathbf{1} \cdot \mathbf{1} = \sum_{i=1}^{m} 1 = m.$$

Nuevo producto escalar en $\mathbb{R}^m$ para la estadística

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_s = \frac{1}{m}(\boldsymbol{x} \cdot \boldsymbol{y})$$

(de manera que: $\|\mathbf{1}\|^2 = \frac{1}{m}\Big(\mathbf{1} \cdot \mathbf{1}\Big) = 1$)

3 La media aritmética

La media aritmética $\mu_{\boldsymbol{y}}$ es el producto escalar de $\boldsymbol{y}$ con $\mathbf{1}$

$$\mu_{\boldsymbol{y}} = \frac{1}{m}\Big(\mathbf{1}\cdot\boldsymbol{y}\Big), \qquad \text{es decir,} \quad \mu_{\boldsymbol{y}} = \frac{1}{m}\sum_{i=1}^{m} y_i$$

La media aritmética $\mu_{\boldsymbol{y}}$ es el *valor* por el que multiplicar $\mathbf{1}$ para obtener la proyección ortogonal de $\boldsymbol{y}$ sobre $\mathcal{L}\big([\mathbf{1};]\big)$

$\boldsymbol{\mu_y}$: proyección de $\boldsymbol{y}\in\mathbb{R}^m$ sobre la recta $\mathcal{L}\big([\mathbf{1};]\big)\subset\mathbb{R}^m$

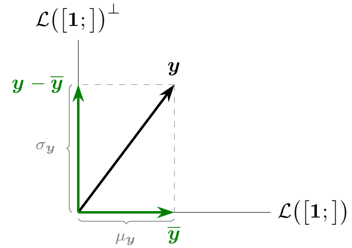
$$\boxed{\boldsymbol{\mu_y} = \mathbf{1}\widehat{a}} \quad \text{y} \quad \boxed{(\boldsymbol{y}-\boldsymbol{\mu_y})\ \perp\ \mathbf{1} \ \Rightarrow\ \tfrac{1}{m}(\boldsymbol{y}-\boldsymbol{\mu_y})\cdot\mathbf{1} = 0}$$

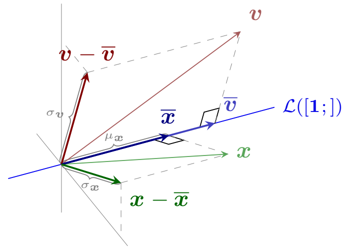
$$\frac{1}{m}(\boldsymbol{y}-\mathbf{1}\widehat{a})\cdot\mathbf{1} = 0 \ \Leftrightarrow\ \frac{1}{m}\Big(\boldsymbol{y}\cdot\mathbf{1}\Big) - \frac{1}{m}\Big(\mathbf{1}\cdot\mathbf{1}\Big)\widehat{a} = 0;$$

Por tanto

$$\widehat{a} = \frac{1}{m}\Big(\boldsymbol{y}\cdot\mathbf{1}\Big) = \mu_{\boldsymbol{y}}$$

4 La media aritmética

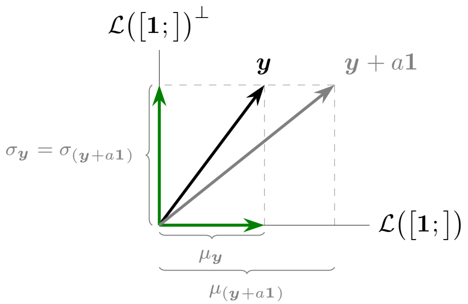




5 Desviación típica

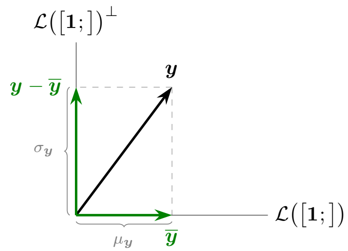
$$\sigma_{\boldsymbol{y}} = \|\boldsymbol{y}-\boldsymbol{\mu_y}\|.$$

Sumar a $\boldsymbol{y}$ un vector constante $a\mathbf{1}$ no cambia la desviación típica.



6 Varianza y el Teorema de Pitagoras

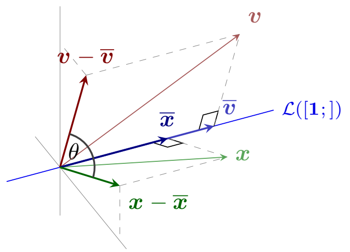
$$\sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y}-\boldsymbol{\mu_y}\|^2 = \frac{1}{m}(\boldsymbol{y}-\boldsymbol{\mu_y})\cdot(\boldsymbol{y}-\boldsymbol{\mu_y}) = \frac{1}{m}\sum_i (y_i-\mu_{\boldsymbol{y}})^2.$$

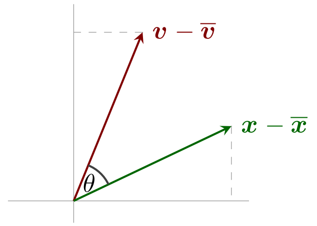


$$\sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y}-\boldsymbol{\mu_y}\|^2 = \|\boldsymbol{y}\|^2 - \|\boldsymbol{\mu_y}\|^2 = \frac{1}{m}\Big(\boldsymbol{y}\cdot\boldsymbol{y}\Big) - \mu_{\boldsymbol{y}}^2, = \frac{\sum_i y_i^2}{m} - \mu_{\boldsymbol{y}}^2.$$

7 Covarianza y correlación

$$\sigma_{\boldsymbol{xy}} = \frac{1}{m}(\boldsymbol{x}-\boldsymbol{\mu_x})\cdot(\boldsymbol{y}-\boldsymbol{\mu_y});$$





$$\rho_{\boldsymbol{xy}} = \frac{\frac{1}{m}(\boldsymbol{x}-\boldsymbol{\mu_x})\cdot(\boldsymbol{y}-\boldsymbol{\mu_y})}{\|(\boldsymbol{x}-\boldsymbol{\mu_x})\|\cdot\|(\boldsymbol{y}-\boldsymbol{\mu_y})\|} = \frac{\sigma_{\boldsymbol{xy}}}{\sqrt{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}}} = \cos(\theta).$$

8 Mínimos Cuadrados Ordinarios (MCO)

Sea $\mathbf{X}$ tal que $\mathcal{L}\big([\mathbf{1};]\big) \subset \mathcal{C}(\mathbf{X})$.

Denotamos con $\widehat{\boldsymbol{y}}$ la proyección ortogonal de $\boldsymbol{y} \in \mathbb{R}^m$ sobre $\mathcal{C}(\mathbf{X})$

$$\boxed{\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}} \quad \text{y} \quad \boxed{(\boldsymbol{y}-\widehat{\boldsymbol{y}}) \perp \mathcal{C}(\mathbf{X}) \;\Rightarrow\; \tfrac{1}{m}\mathbf{X}^{\intercal}(\boldsymbol{y}-\widehat{\boldsymbol{y}}) = \mathbf{0}}$$

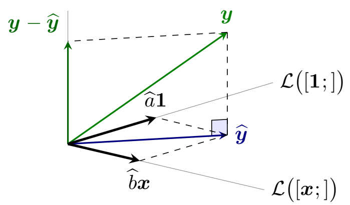
$$\frac{1}{m}\mathbf{X}^{\intercal}(\boldsymbol{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0} \iff \frac{1}{m}\mathbf{X}^{\intercal}\boldsymbol{y} - \frac{1}{m}\mathbf{X}^{\intercal}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}.$$

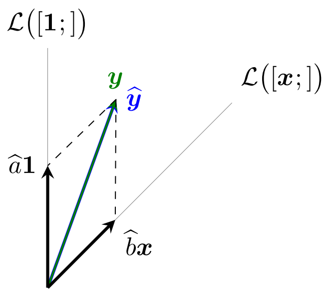
Por tanto

$$\Big(\frac{1}{m}\mathbf{X}^{\intercal}\mathbf{X}\Big)\widehat{\boldsymbol{\beta}} = \frac{1}{m}\mathbf{X}^{\intercal}\boldsymbol{y}.$$

9 Mínimos Cuadrados Ordinarios (MCO)

Si $\mathbf{X} = \big[\mathbf{1};\, \boldsymbol{x};\big]$ es de rango 2.

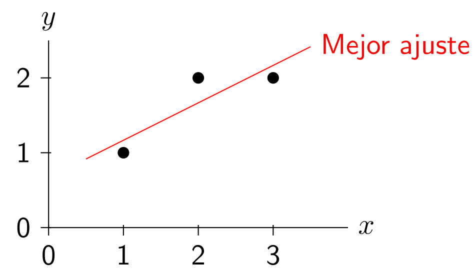




$$\Big(\frac{1}{m}\mathbf{X}^{\intercal}\mathbf{X}\Big)\begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} = \frac{1}{m}\mathbf{X}^{\intercal}\boldsymbol{y}.$$

10 Aplicación: Mínimos cuadrados (Ajuste lineal)

"buscando la mejor recta de ajuste $\widehat{y} = \widehat{a} + \widehat{b}x$"
Puntos $(x, y,)$: $(1,1,)$; $(2,2,)$; $(3,2,)$



$$\begin{cases} a+1b &= 1\\ a+2b &= 2\\ a+3b &= 2\end{cases} \quad\rightarrow\quad \begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}\begin{pmatrix}a\\b\end{pmatrix} = \begin{pmatrix}1\\2\\2\end{pmatrix} \qquad (\mathbf{X}\boldsymbol{\beta} = \boldsymbol{y} \text{ Sin solución})$$

11 Aplicación: Mínimos cuadrados (Ajuste lineal)

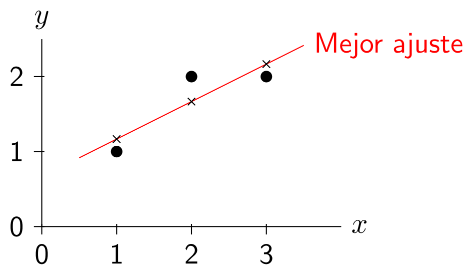
$$\mathbf{X}\beta = \boldsymbol{y} \quad \text{(Sin solución)} \;\rightarrow\; \left(\frac{1}{m}\mathbf{X}^\intercal\mathbf{X}\right)\widehat{\beta} = \frac{1}{m}\mathbf{X}^\intercal\boldsymbol{y} \quad\rightarrow\quad \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

$$\begin{bmatrix}1&1&1\\1&2&3\end{bmatrix}\begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}\begin{pmatrix}\widehat{a}\\\widehat{b}\end{pmatrix} = \begin{bmatrix}1&1&1\\1&2&3\end{bmatrix}\begin{pmatrix}1\\2\\2\end{pmatrix}$$

$$\begin{bmatrix}3&6\\6&14\end{bmatrix}\begin{pmatrix}\widehat{a}\\\widehat{b}\end{pmatrix} = \begin{pmatrix}5\\11\end{pmatrix} \quad\Rightarrow\quad \widehat{a} = \frac{2}{3};\quad \widehat{b} = \frac{1}{2}.$$

$$\boxed{\text{Mejor solución: } \tfrac{2}{3} + \tfrac{1}{2}x}$$

12 Aplicación: Mínimos cuadrados (Ajuste lineal)

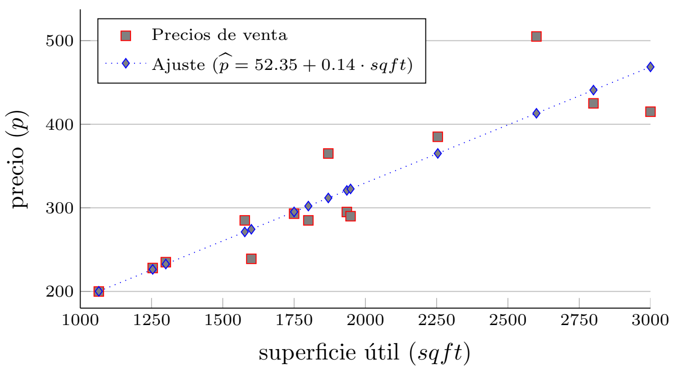


$$\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}\begin{pmatrix}\widehat{a}\\\widehat{b}\end{pmatrix} \longrightarrow \widehat{\boldsymbol{y}} = \begin{pmatrix}7/6\\10/6\\13/6\end{pmatrix} \longrightarrow \widehat{\boldsymbol{e}} = \begin{pmatrix}-1/6\\2/6\\-1/6\end{pmatrix}$$

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \quad \text{y} \quad \begin{cases}\widehat{\boldsymbol{e}}\cdot\widehat{\boldsymbol{y}} &= 0\\ \widehat{\boldsymbol{e}}\mathbf{X} &= \mathbf{0}\end{cases}.$$

13 Aplicación: ajustando por mínimos cuadrados

Precios de venta (miles de dólares) y superficie útil (pies al cuadrado) de 14 casas unifamiliares en la *University City* de la ciudad de San Diego en California en 1990 (Ramanathan, 2002, pp. 78)



Problemas de la Lección 19

(L-19) PROBLEMA 1. Con las medidas $\boldsymbol{y} = (0, 8, 8, 20,)$ tomadas en los instantes $\boldsymbol{x} = (0, 1, 3, 4,)$,

(a) Plantee y resuelva las ecuaciones normales $\mathbf{A}^\intercal\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}^\intercal\boldsymbol{y}$.
(b) Para el mejor ajuste lineal, encuentre los ajustes p_i y los cuatro errores e_i.
(c) ¿Cuál es el cuadrado de la norma del vector de errores $\|\boldsymbol{e}\|^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2$?
(d) Dibuje la recta de regresión
(e) Sustituya las medidas $\boldsymbol{y}$ por los valores ajustados $\boldsymbol{p} = (1, 5, 13, 17,)$ escriba las cuatro ecuaciones $\mathbf{A}\boldsymbol{\beta} = \boldsymbol{p}$. Encuentre la solución exacta a $\mathbf{A}\boldsymbol{\beta} = \boldsymbol{p}$
(f) Verifique que $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{p} = (-1, 3, -5, 3,)$ es perpendicular a las dos columnas de $\mathbf{A}$.
(g) ¿Cuál es la distancia más corta $\|\boldsymbol{e}\|$ desde $\boldsymbol{y}$ al espacio columna de $\mathbf{A}$?

(Strang, 2003, ejercicio 1–3 del conjunto de problemas 4.3.)

(L-19) PROBLEMA 2.

(a) Escriba las tres ecuaciones $y = \alpha + \beta x$ dado el conjunto de datos: $y = 7$ para $x = -1$, $y = 7$ para $x = 1$, y $y = 21$ para $x = 2$. Encuentre la solución de mínimos cuadrados $\widehat{\boldsymbol{\beta}} = (\hat{\alpha}, \hat{\beta})$ y pinte el mejor ajuste lineal.
(b) Encuentre la proyección $\boldsymbol{p} = \mathbf{A}\widehat{\boldsymbol{\beta}}$. Es decir, tres valores del mejor ajuste lineal. Demuestre que el vector de error es $\boldsymbol{e} = (2, -6, 4,)$. ¿Por que es $\mathbf{P}\boldsymbol{e} = \mathbf{0}$?

(L-19) PROBLEMA 3. Our measurements at times $t = 1, 2, 3$ are $b = 1, 4$, and b_3. We want to fit those points by the nearest line $C + Dt$, using least squares.

(a) Which value for b_3 will put the three measurements on a straight line? *Which line is it?* Will least squares choose that line if the third measurement is $b_3 = 9$? (Yes or no).

(b) What is the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ that would be solved exactly for $\boldsymbol{x} = (C, D)$ if the three points do lie on a line? Compute the projection matrix $\mathbf{P}$ onto the column space of $\mathbf{A}$.

(c) What is the rank of that projection matrix $\mathbf{P}$? How is the column space of $\mathbf{P}$ related to the column space of $\mathbf{A}$? (You can answer with or without the entries of $\mathbf{P}$ computed in (b).)

(d) Suppose $b_3 = 1$. Write down the equation for the best least squares solution $\widehat{\boldsymbol{x}}$, and show that the best straight line is horizontal.

Ramanathan, R. (2002). *Introductory Econometrics with applications*. South-Western, Mason, Ohio, fifth ed. ISBN 0-03-034186-8.

Strang, G. (2003). *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, Massachusetts. USA, third ed. ISBN 0-9614088-9-8.