Mathematics II

Marcos Bujosa

Universidad Complutense de Madrid

14/01/2023

1 Highlights of **Lesson 19**

Highlights of Lesson 19

- Mean
- Standard deviation and variance
- Ordinary Least Squares (OLS)

2 Restriction in statistics and probability

> Norm of constant vector "one" is 1

This fails using the dot product in $\mathbb{R}^m$ $(m > 1)$

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

New scalar product in $\mathbb{R}^m$ for statistics

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

(so: $\|\mathbf{1}\|^2 = \frac{1}{m}\left(\mathbf{1} \cdot \mathbf{1}\right) = 1$)

$\boxed{3}$ Mean

The mean $\mu_{\boldsymbol{y}}$ is the scalar product of $\boldsymbol{y}$ and $\mathbf{1}$

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

The mean $\mu_{\boldsymbol{y}}$ is the *value* by which to multiply $\mathbf{1}$ to get the orthogonal projection of $\boldsymbol{y}$ onto $\mathcal{L}\big([\mathbf{1};]\big)$

$\mu_{\boldsymbol{y}}$: projection of $\boldsymbol{y} \in \mathbb{R}^m$ onto the line $\mathcal{L}\big([\mathbf{1};]\big) \subset \mathbb{R}^m$

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

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

Therefore

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

$\boxed{4}$ Mean
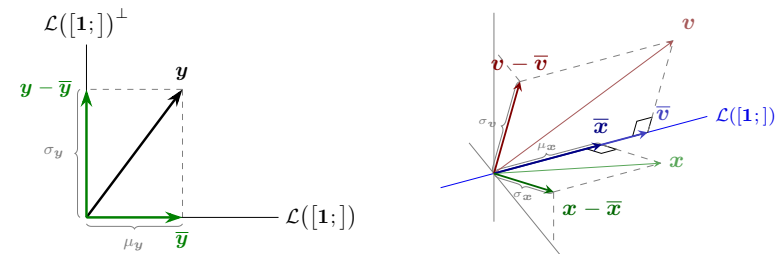
$\boxed{5}$ Standard deviation

$$\sigma_{\boldsymbol{y}} = \|\boldsymbol{y} - \mu_{\boldsymbol{y}}\|.$$
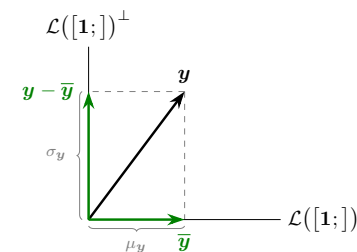
Adding a constant vector $a\mathbf{1}$ to $\boldsymbol{y}$ does not change the standard deviation.
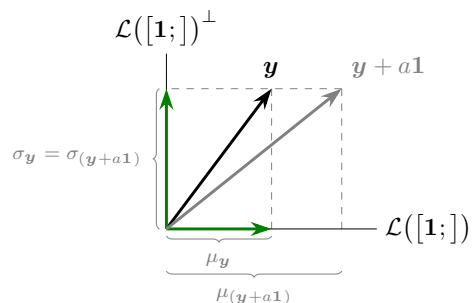
$\boxed{6}$ Variance and the Pythagorean theorem

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

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

7 Covariance and correlation

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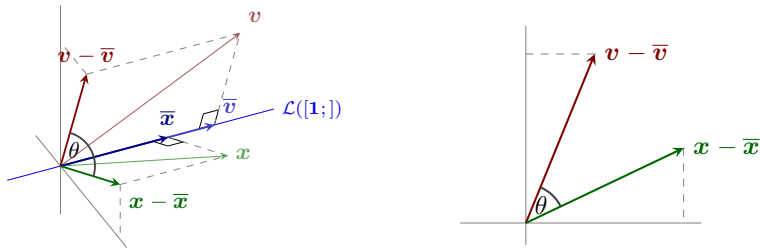
$$\rho_{xy} = \frac{\frac{1}{m}(x - \mu_x) \cdot (y - \mu_y)}{\|(x - \mu_x)\| \cdot \|(y - \mu_y)\|} = \frac{\sigma_{xy}}{\sqrt{\sigma_x \sigma_y}} = \cos(\theta).$$

8 Ordinary Least Squares (OLS)

Let $\mathbf{X}$ suach that $\mathcal{L}\big([\mathbf{1};]\big) \subset \mathcal{C}(\mathbf{X})$.

$\widehat{y}$ is the orthogonal projection of $y \in \mathbb{R}^m$ onto $\mathcal{C}(\mathbf{X})$

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

$$\frac{1}{m}\mathbf{X}^{\mathsf{T}}(y - \mathbf{X}\widehat{\beta}) = \mathbf{0} \quad \Longleftrightarrow \quad \frac{1}{m}\mathbf{X}^{\mathsf{T}}y - \frac{1}{m}\mathbf{X}^{\mathsf{T}}\mathbf{X}\widehat{\beta} = \mathbf{0}.$$

Therefore

$$\left(\frac{1}{m}\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)\widehat{\beta} = \frac{1}{m}\mathbf{X}^{\mathsf{T}}y.$$

9 Ordinary Least Squares (OLS)
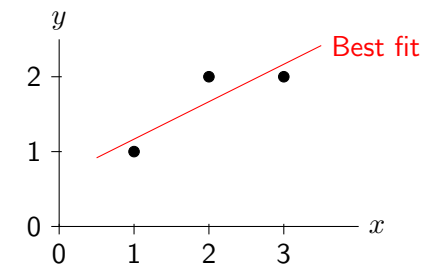
If $\mathbf{X} = \big[\mathbf{1}; x;\big]$ has rank 2.
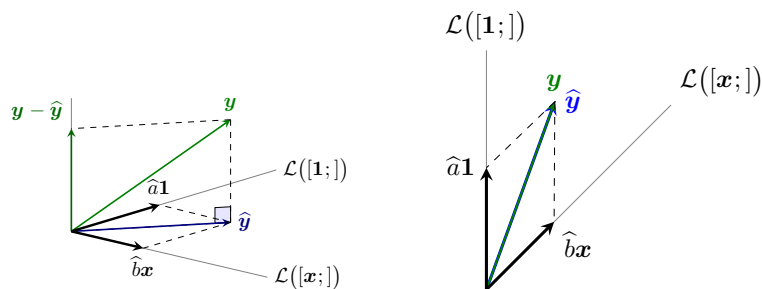
$$\left(\frac{1}{m}\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)\begin{pmatrix}\widehat{a}\\\widehat{b}\end{pmatrix} = \frac{1}{m}\mathbf{X}^{\mathsf{T}}y.$$

10 Application: Least Squares (Fitting by a line)

"looking for the best fitting line $\widehat{y} = \widehat{a} + \widehat{b}x$"
Points $(x, y,)$: $(1, 1,)$; $(2, 2,)$; $(3, 2,)$

$$\begin{cases} a + 1b &= 1 \\ a + 2b &= 2 \\ a + 3b &= 2 \end{cases} \quad \rightarrow \quad \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad (\mathbf{X}\beta = y \text{ No solution})$$

11 Application: Least Squares (Fitting by a line)

$$\mathbf{X}\boldsymbol{\beta} = \boldsymbol{y} \quad (\text{No solution}) \;\to\; \left(\frac{1}{m}\mathbf{X}^\mathsf{T}\mathbf{X}\right)\widehat{\boldsymbol{\beta}} = \frac{1}{m}\mathbf{X}^\mathsf{T}\boldsymbol{y} \;\to\; \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}.$$
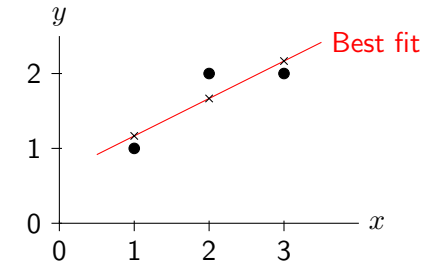
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

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

Best solution: $\frac{2}{3} + \frac{1}{2}x$

12 Application: Least Squares (Fitting by a line)

Best fit

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

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \quad \text{and} \quad \begin{cases} \widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{y}} &= 0 \\ \widehat{\boldsymbol{e}}\mathbf{X} &= \mathbf{0} \end{cases}.$$
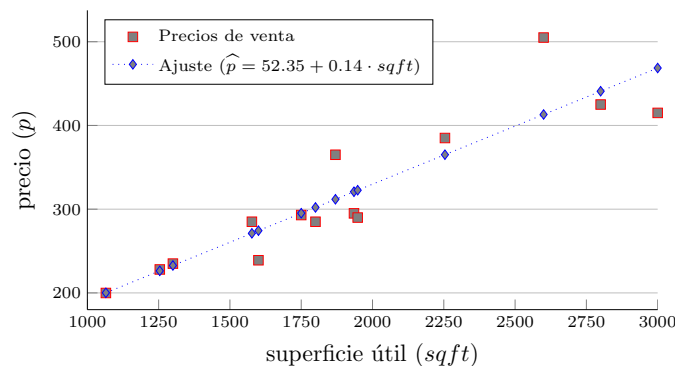
13 Application: Least Squares (Fitting by a line)

Selling price and living area of single family homes in University City community of San Diego, in 1990.

price = Sale price is in thousands of dollars

sqft = Square feet of living area (Ramanathan, 2002, pp. 78)

Questions of the Lecture 19

(L-19) QUESTION 1. With the measurements $\boldsymbol{y} = (0, 8, 8, 20,)$ at $\boldsymbol{x} = (0, 1, 3, 4,)$,

(a) Set up and solve the normal equations $\mathbf{A}^\mathsf{T}\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}^\mathsf{T}\boldsymbol{y}$.
(b) For the best straight line, find its four fits p_i and four errors e_i.
(c) What is the value of the square of the norm of the error vector
$\|\boldsymbol{e}\|^2 = e_1^2 + e_2^2 + e_3^2 + e_4^2$?
(d) Draw the regression line
(e) Change the measurements to $\boldsymbol{p} = (1, 5, 13, 17,)$ write down the four equations
$\mathbf{A}\boldsymbol{\beta} = \boldsymbol{p}$. Find an exact solution to $\mathbf{A}\boldsymbol{\beta} = \boldsymbol{p}$
(f) Check that $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{p} = (-1, 3, -5, 3,)$ is perpendicular to both columns of the
same matrix $\mathbf{A}$.
(g) What is the shortest distance $\|\boldsymbol{e}\|$ from $\boldsymbol{y}$ to the column space of $\mathbf{A}$?

(Strang, 2003, exercise 1–3 from section 4.3.)

(L-19) QUESTION 2.

(a) Write down three equations $y = \alpha + \beta x$ given the data: $y = 7$ at $x = -1$, $y = 7$
at $x = 1$, and $y = 21$ at $x = 2$. Find the least squares solution $\widehat{\boldsymbol{\beta}} = (\widehat{\alpha}, \widehat{\beta})$ and
draw the closest line.
(b) Find the projection $\boldsymbol{p} = \mathbf{A}\widehat{\boldsymbol{\beta}}$. This gives the three heights of the closest line. Show
that the error vector is $\boldsymbol{e} = (2, -6, 4,)$. Why is $\mathbf{P}\boldsymbol{e} = \mathbf{0}$?

(L-19) QUESTION 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.