Mathematics II

Marcos Bujosa

03/05/2023

You can find the last version of these course materials at

https://github.com/mbujosab/MatematicasII/tree/main/Eng

Contents

Part VII
Geometrical view of statistics

LECTURE 19: Statistics

$\boxed{\textit{Lecture 19}}$

(**Lecture** 19) $\boxed{\text{S-1}}$ Highlights of Lesson 19

Highlights of *Lesson 19*

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

F1

(**Lecture** 19) $\boxed{\text{S-2}}$ Restriction in statistics and probability

$\boxed{\text{Norm of constant vector “one” is 1}}$

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

F2

(**Lecture** 19) $\boxed{\text{S-3}}$ 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 $\mathcal{L}\big([\mathbf{1};]\big)$

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

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

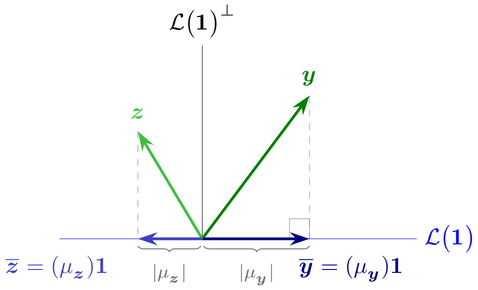
$$\frac{1}{m}(\boldsymbol{y} - \mathbf{1}\widehat{a}) \cdot \mathbf{1} = 0 \Leftrightarrow \frac{1}{m}\Big(\boldsymbol{y} \cdot \mathbf{1}\Big) - \frac{1}{m}\Big(\mathbf{1} \cdot \mathbf{1}\Big)\widehat{a} = 0;$$

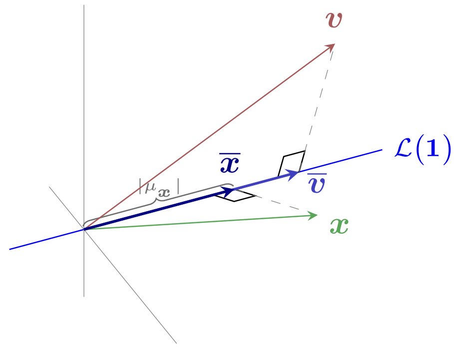
Therefore

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

F3

(**Lecture** 19) S-4 Mean

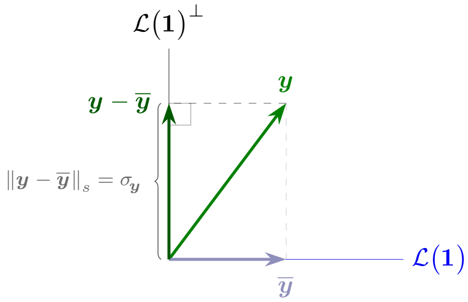


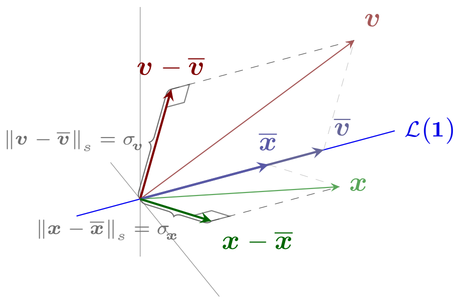


F4

(**Lecture** 19) S-5 Standard deviation

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

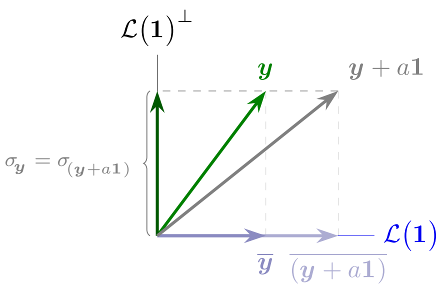


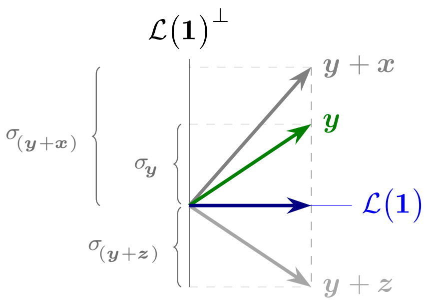


F5

(**Lecture** 19) S-6 Constant Vectors and Zero Mean Vectors

Adding a constant vector $a\mathbf{1}$ to $\boldsymbol{y}$ does not change the standard deviation.



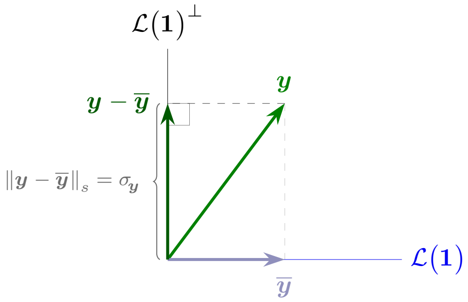


$$\sigma_{\boldsymbol{z}} = 0 \;\Leftrightarrow\; \boldsymbol{z} = a\mathbf{1}; \qquad \mu_{\boldsymbol{z}} = 0 \;\Leftrightarrow\; \boldsymbol{z} \perp \mathbf{1}$$

F6

(**Lecture** 19) S-7 Variance and the Pythagorean theorem

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

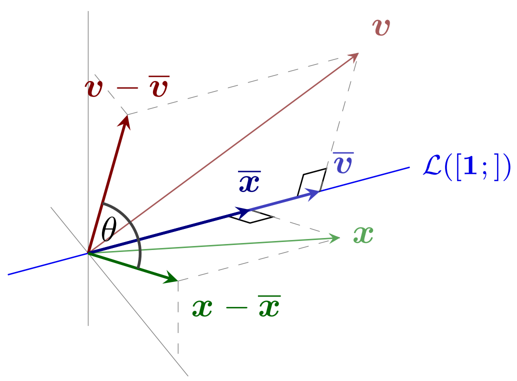


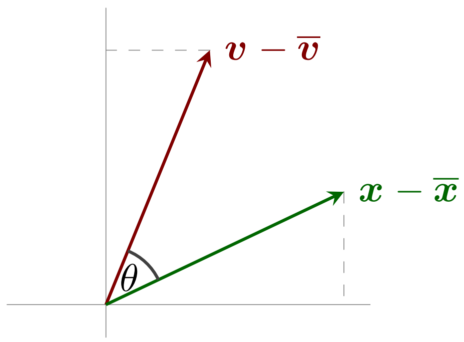
$$\sigma_{\boldsymbol{y}}^2 = \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|^2 = \|\boldsymbol{y}\|^2 - \|\overline{\boldsymbol{y}}\|^2 = \frac{1}{m}\Big(\boldsymbol{y}\cdot\boldsymbol{y}\Big) - \mu_{\boldsymbol{y}}^2, = \frac{\sum_i y_i^2}{m} - \mu_{\boldsymbol{y}}^2.$$

F7

(**Lecture** 19) S-8 Covariance and correlation

$$\sigma_{\boldsymbol{x}\boldsymbol{y}} = \frac{1}{m}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\cdot(\boldsymbol{y}-\overline{\boldsymbol{y}});$$





$$\rho_{\boldsymbol{x}\boldsymbol{y}} = \frac{\frac{1}{m}(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\cdot(\boldsymbol{y}-\overline{\boldsymbol{y}})}{\|(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}})\|\cdot\|(\boldsymbol{y}-\overline{\boldsymbol{y}})\|} = \frac{\sigma_{\boldsymbol{x}\boldsymbol{y}}}{\sqrt{\sigma_{\boldsymbol{x}}\sigma_{\boldsymbol{y}}}} = \cos(\theta).$$

F8

(**Lecture** 19) S-9 **Ordinary Least Squares (OLS)**

Let $\mathbf{X}$ suach that $\mathcal{L}([\mathbf{1};]) \subset \mathcal{C}(\mathbf{X})$.
$\widehat{\boldsymbol{y}}$ is the orthogonal projection of $\boldsymbol{y} \in \mathbb{R}^m$ onto $\mathcal{C}(\mathbf{X})$

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

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

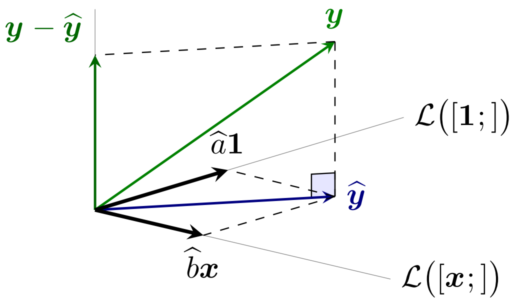
Therefore

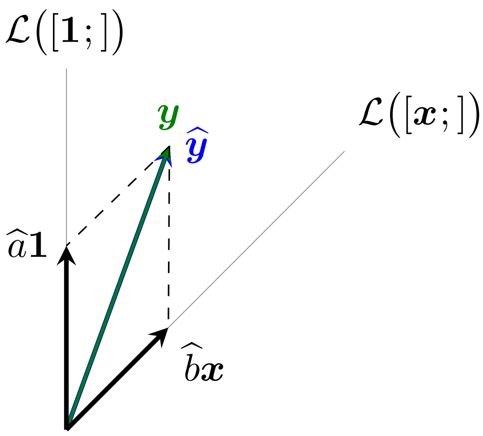
$$\left(\frac{1}{m}\mathbf{X}^\intercal\mathbf{X}\right)\widehat{\boldsymbol{\beta}} = \frac{1}{m}\mathbf{X}^\intercal\boldsymbol{y}.$$

F9

(**Lecture** 19) S-10 **Ordinary Least Squares (OLS)**

If $\mathbf{X} = [\mathbf{1}; \boldsymbol{x};]$ has rank 2.



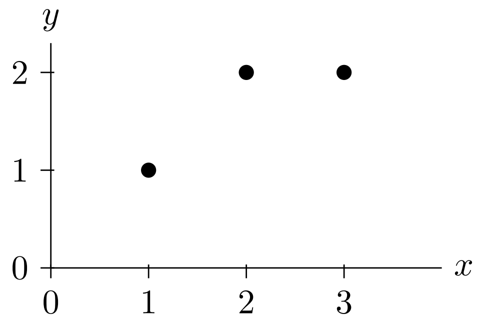


$$\left(\frac{1}{m}\mathbf{X}^\intercal\mathbf{X}\right)\begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} = \frac{1}{m}\mathbf{X}^\intercal\boldsymbol{y}.$$

F10

(**Lecture** 19) S-11 **Application: Least Squares (Fitting by a line)**

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



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

F11

(**Lecture** 19) S-12 **Application: Least Squares (Fitting by a line)**

$$\mathbf{X}\boldsymbol{\beta} = \boldsymbol{y} \quad (\text{No solution}) \;\to\; \left(\frac{1}{m}\mathbf{X}^\intercal\mathbf{X}\right)\widehat{\boldsymbol{\beta}} = \frac{1}{m}\mathbf{X}^\intercal\boldsymbol{y} \quad \to \quad \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

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

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

$$\boxed{\text{Best solution: } \tfrac{2}{3} + \tfrac{1}{2}x}$$

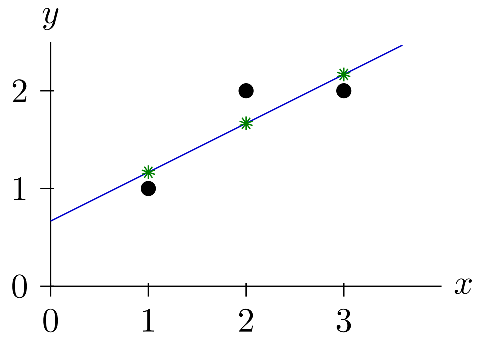
F12

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

and

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \begin{pmatrix} 7/6 \\ 10/6 \\ 13/6 \end{pmatrix} = \begin{pmatrix} -1/6 \\ 2/6 \\ -1/6 \end{pmatrix}$$

(**Lecture** 19) S-13 Application: Least Squares (Fitting by a line)



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

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

F13

$$\widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{p}} = \begin{pmatrix} -1/6, & 2/6, & -1/6, \end{pmatrix} \cdot \begin{pmatrix} 7/6 \\ 10/6 \\ 13/6 \end{pmatrix} = 0; \qquad \widehat{\boldsymbol{e}}\mathbf{A} = \begin{pmatrix} -1/6, & 2/6, & -1/6, \end{pmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{pmatrix} 0, & 0, \end{pmatrix}$$

Example 1. [**Selling price and living area of single family homes:**]
Consider the following data. Selling price and living area of single family homes in University City community of

n	Price ($\boldsymbol{y}$)	Sqrt ($\boldsymbol{x}$)
1	199.9	1065
2	228.0	1254
3	235.0	1300
4	285.0	1577
5	239.0	1600
6	293.0	1750
7	285.0	1800
8	365.0	1870
9	295.0	1935
10	290.0	1948
11	385.0	2254
12	505.0	2600
13	425.0	2800
14	415.0	3000

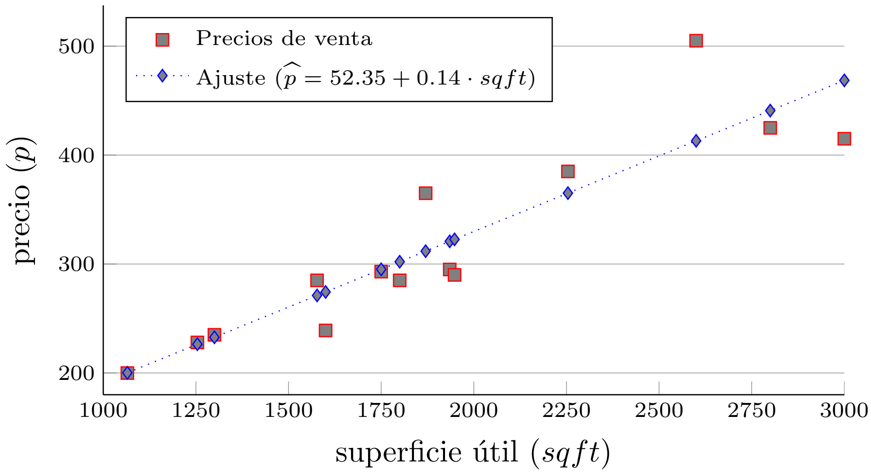
San Diego, in 1990.
price = Sale price is in thousands of dollars
sqft = Square feet of living area(**?**, pp. 78)

(**Lecture** 19) S-14 **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 (**?**, pp. 78)



F14

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

n	Price	living area	Fitted price	Error $\widehat{e}$
1	199.9	1065	200.1200	-0.22000
2	228.0	1254	226.3438	1.65619
3	235.0	1300	232.7263	2.27368
4	285.0	1577	271.1602	13.83984
5	239.0	1600	274.3514	-35.35142
6	293.0	1750	295.1640	-2.16397
7	285.0	1800	302.1015	-17.10148
8	365.0	1870	311.8140	53.18600
9	295.0	1935	320.8328	-25.83278
10	290.0	1948	322.6365	-32.63653
11	385.0	2254	365.0941	19.90587
12	505.0	2600	413.1017	91.89826
13	425.0	2800	440.8518	-15.85180
14	415.0	3000	468.6019	-53.60187

price = Sale price is in thousands of dollars

sqft = Square feet of living area (**?**, pp. 78)

The lecture ends here

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 $\mathbf{A}$?

(**?**, 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 $\widehat{\boldsymbol{\beta}} = (\hat{\alpha}, \hat{\beta})$ 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.

MIT 18.06 - Quiz 2, November 2, 2005

End of Questions of the Lecture 19

Solutions

(L-19) Question 1(a) For the $\boldsymbol{y}$ and the given points our matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} 0 \\ 8 \\ 8 \\ 20 \end{pmatrix}$$

The normal equations are given by $\mathbf{A}^\mathsf{T}\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}^\mathsf{T}\boldsymbol{y}$, or

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{bmatrix} \begin{pmatrix} 0 \\ 8 \\ 8 \\ 20 \end{pmatrix}$$

or

$$\begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} 36 \\ 112 \end{pmatrix}$$

which has as its solution $\widehat{\alpha} = 1$, $\widehat{\beta} = 4$.

$\square$

(L-19) Question 1(b) So the four heights with this $\widehat{\boldsymbol{\beta}}$ are given by

$$\boldsymbol{p} = \mathbf{A}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 13 \\ 17 \end{pmatrix}.$$

With this solution by direct calculation the error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is given by

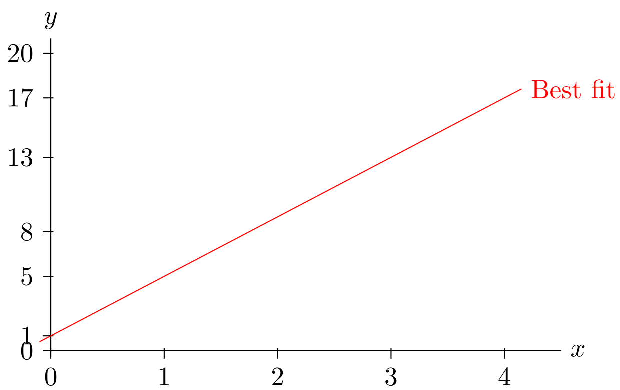
$$\boldsymbol{e} = \begin{pmatrix} 0 \\ 8 \\ 8 \\ 20 \end{pmatrix} - \begin{pmatrix} 1 \\ 5 \\ 13 \\ 17 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -5 \\ 3 \end{pmatrix}.$$

$\square$

(L-19) Question 1(c) It is the smallest possible value for a linear fitting $\|\boldsymbol{e}\|^2 = \boldsymbol{e}\cdot\boldsymbol{e} = (-1)^2+3^2+(-5)^2+3^2 = 44$.

$\square$

(L-19) Question 1(d)



$\square$

(L-19) Question 1(e) If our mathematical model of the relationship between $\boldsymbol{y}$ and $\boldsymbol{x}$ is a line given by $\boldsymbol{y} = \alpha\cdot\mathbf{1}+\beta\cdot\boldsymbol{x}$, then if the measurements change to what is given in the text then we have

$$\begin{aligned}\alpha + 0\beta =& 1\\ \alpha + 1\beta =& 5\\ \alpha + 3\beta =& 13\\ \alpha + 4\beta =& 17\end{aligned}$$

Which has as an analytic solution given by $\alpha = 1$ and $\beta = 4$. $\square$

(L-19) Question 1(f)

$$\boldsymbol{e}\mathbf{A} = \begin{pmatrix}-1, & 3, & -5, & 3,\end{pmatrix}\begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 3\\ 1 & 4\end{bmatrix} = \begin{pmatrix}0, & 0,\end{pmatrix}.$$

$\square$

(L-19) Question 1(g) So the shortest distance is given by $\|\boldsymbol{e}\| = \sqrt{\boldsymbol{e}\cdot\boldsymbol{e}} = \sqrt{44}$. $\square$

(L-19) Question 2(a) Our equations are given by

$$\begin{aligned}\alpha - \beta =& 7\\ \alpha + \beta =& 7\\ \alpha + 2\beta =& 21\end{aligned}$$

Which as a system of linear equations matrix are given by

$$\begin{bmatrix}1 & -1\\ 1 & 1\\ 1 & 2\end{bmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}7\\ 7\\ 21\end{pmatrix}$$

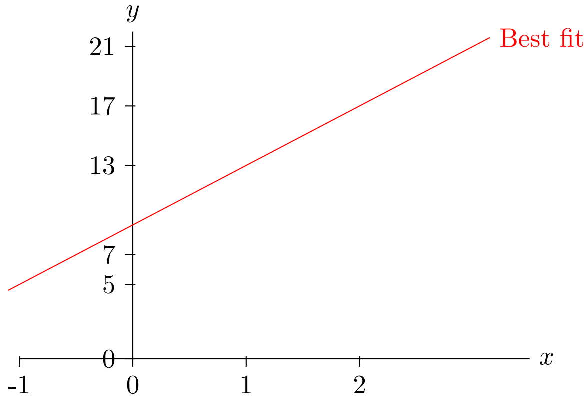
The least squares solution is given by $\mathbf{A}^\intercal\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}^\intercal\boldsymbol{y}$ which in this case simplify as follows

$$\begin{bmatrix}1 & 1 & 1\\ -1 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & -1\\ 1 & 1\\ 1 & 2\end{bmatrix}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} = \begin{bmatrix}1 & 1 & 1\\ -1 & 1 & 2\end{bmatrix}\begin{pmatrix}7\\ 7\\ 21\end{pmatrix}$$

or

$$\begin{bmatrix}3 & 2\\ 2 & 6\end{bmatrix}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} = \begin{pmatrix}35\\ 42\end{pmatrix}$$

which gives the following $\widehat{\alpha} = 9$, $\widehat{\beta} = 4$. So the best linear fit is $\widehat{y} = 9 + 4x$.



□

(L-19) Question 2(b)

$$\boldsymbol{p} = \mathbf{A}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} 9 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 \\ 13 \\ 17 \end{pmatrix}.$$

that gives the values on the closest line. The error vector $\boldsymbol{e}$ is then given by

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{p} = \begin{pmatrix} 7 \\ 7 \\ 21 \end{pmatrix} - \begin{pmatrix} 5 \\ 13 \\ 17 \end{pmatrix} = \begin{pmatrix} 2 \\ -6 \\ 4 \end{pmatrix}.$$

The matrix $\mathbf{P}$ projects onto the column space of $\mathbf{A}$, but $\boldsymbol{e}$ is in the left null space, so it is orthogonal to $\mathcal{C}\left(\mathbf{A}\right)$.

□

(L-19) Question 3(a) The three data points lie on the same line when $b_3 = 7$. This line is $-2 + 3t$. If $b_3 = 9$, the least squares method will NOT choose this line. (A quick way to see this is from the fact that the line chosen by least squares will give the average of the given b's at the time equal to the average of the given t's; in this case, the best fit line would take the value $(1 + 3 + 9)/3 = 13/3$ at $t = (1 + 2 + 3)/3 = 2$, whereas our line gives 4 at $t = 2$.)

□

(L-19) Question 3(b) The linear system for $\boldsymbol{x} = (C, D)$ would be the following:

$$\mathbf{A}\boldsymbol{x} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ b_3 \end{pmatrix}.$$

We compute the projection matrix $\mathbf{P}$ onto the column space of $\mathbf{A}$ using the projection matrix formula:

$$\begin{aligned} \mathbf{P} =& \mathbf{A}(\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}} = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \\ =& \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 8 & 4 & -4 \\ -3 & 0 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}. \end{aligned}$$

□

(L-19) Question 3(c) The column space of $\mathbf{P}$ is the space consisting of all the vectors $\mathbf{P}\boldsymbol{b}$, i.e. all the projections of vectors in $\mathbb{R}^3$ onto the column space of $\mathbf{A}$, which is precisely the column space of $\mathbf{A}$. Thus the rank of $\mathbf{P}$ is equal to the rank of $\mathbf{A}$, which is 2.

□

(L-19) Question 3(d) The equation for the best least squares solution $\widehat{\boldsymbol{x}}$ is $\mathbf{A}^{\mathsf{T}}\mathbf{A}\widehat{\boldsymbol{x}} = \mathbf{A}^{\mathsf{T}}\boldsymbol{b}$, where $\boldsymbol{b} = \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}$. Writing out this system, we get

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \widehat{\boldsymbol{x}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \end{pmatrix}.$$

The solution to this system is $\widehat{\boldsymbol{x}} = \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, so the best fit line is the horizontal line $b = 2$.

□