Today

Least squares
Least squares model fitting

How do we find the best line to fit through the data?
Least squares model fitting

Each red bar is called a residual. We want all the residuals to be as small as possible.
The residuals are...

(A) $r_i = y_i^2 + x_i^2$

(B) $r_i = a^2 (y_i^2 + x_i^2)$

(C) $r_i = y_i - ax_i$

(D) $r_i = y_i - x_i$

(E) $r_i = x_i - y_i$
The residuals are...

(A) \( r_i = y_i^2 + x_i^2 \)
(B) \( r_i = a^2 (y_i^2 + x_i^2) \)
(C) \( r_i = y_i - ax_i \)
(D) \( r_i = y_i - x_i \)
(E) \( r_i = x_i - y_i \)
To minimize the residuals, we define the objective function...

(A) \[ f(a) = |y_1-ax_1| + |y_2-ax_2| + \ldots + |y_n-ax_n| \]

(B) \[ f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 + \ldots + (y_n-ax_n)^2 \]

(C) \[ f(a) = (y_1-ax_1)(y_2-ax_2)\ldots(y_n-ax_n) \]

(D) \[ f(a) = (ay_1-x_1)^2 + (ay_2-x_2)^2 + \ldots + (ay_n-x_n)^2 \]
To minimize the residuals, we define the objective function...

(A) \( f(a) = |y_1-ax_1| + |y_2-ax_2| + \ldots + |y_n-ax_n| \)

(B) \( f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 + \ldots + (y_n-ax_n)^2 \)

(C) \( f(a) = (y_1-ax_1)(y_2-ax_2)\ldots(y_n-ax_n) \)

(D) \( f(a) = (ay_1-x_1)^2 + (ay_2-x_2)^2 + \ldots + (ay_n-x_n)^2 \)

(B) is called the “sum of squared residuals”. (A) is also reasonable but not as “good” (take a stats class to find out more).
Find a so that \( y=ax \) fits (4,5), (6,7) in the “least squares” sense.

Define \( f(a) \):

(A) \( \text{SSR}(a) = |5-4a| + |7-6a| \)

(B) \( \text{SSR}(a) = (4-5a)^2 + (6-7a)^2 \)

(C) \( \text{SSR}(a) = (5-4a)^2 + (7-6a)^2 \)

(D) \( \text{SSR}(a) = (5-4-a)^2 + (7-6-a)^2 \)
Find a so that \( y=ax \) fits \((4,5), (6,7)\) in the “least squares” sense.

Define \( f(a) \):

(A) \( \text{SSR}(a) = |5-4a| + |7-6a| \)

(B) \( \text{SSR}(a) = (4-5a)^2 + (6-7a)^2 \)

(C) \( \text{SSR}(a) = (5-4a)^2 + (7-6a)^2 \)

(D) \( \text{SSR}(a) = (5-4-a)^2 + (7-6-a)^2 \)

Recall: \( f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2 \)
Find a so that $y=ax$ fits (4,5), (6,7) in the “least squares” sense.

Find the $a$ that minimizes $SSR(a)$:

(A) $a = \frac{7}{6}$

(B) $a = \frac{5}{4}$

(C) $a = \frac{\left(\frac{7}{6} + \frac{5}{4}\right)}{2}$

(D) $a = \frac{31}{26}$
Find a so that $y=ax$ fits (4,5), (6,7) in the “least squares” sense.

Find the $a$ that minimizes $SSR(a)$:

$$SSR(a) = (5-4a)^2 + (7-6a)^2$$
$$= 5^2 -2\cdot 4\cdot 5a + 4^2a^2 + 7^2 -2\cdot 6\cdot 7a + 6^2a^2$$

$$SSR'(a) = -2\cdot 4\cdot 5 + 2\cdot 4^2a -2\cdot 6\cdot 7 + 2\cdot 6^2a = 0$$

$$a = \frac{2\cdot 4\cdot 5 + 2\cdot 6\cdot 7}{2\cdot 4^2 + 2\cdot 6^2}$$
$$= \frac{4\cdot 5 + 6\cdot 7}{4^2 + 6^2} = \frac{62}{52}$$
$$= \frac{x_1\cdot y_1 + x_2\cdot y_2}{x_1^2 + x_2^2}$$

Desmos + mean + $\ell^1_1$ on doc cam (if time).
A model is a function that you use to summarize or fit data. For example, some common ones: \( f(x) = ax, \quad f(x) = ax + b, \quad f(x) = Ce^{-kx}. \)

Residuals are a measure of how far each model value is from the data value: \( r_i = y_i - f(x_i). \)

The Sum of Squared Residuals (SSR) is a measure of how well the model fits all the data: \( SSR = \sum (y_i - f(x_i))^2. \) Small is better.
Spreadsheet warnings

Excel and other spreadsheet don’t give $y=ax$ fits, just $y=ax+b$. Don’t use a $y=ax+b$ function to get $a$ for $y=ax$.

Trendline in Excel’s chart gives $y=ax+b$ and usually not to enough decimals for WW.

Do these “manually” using cells.