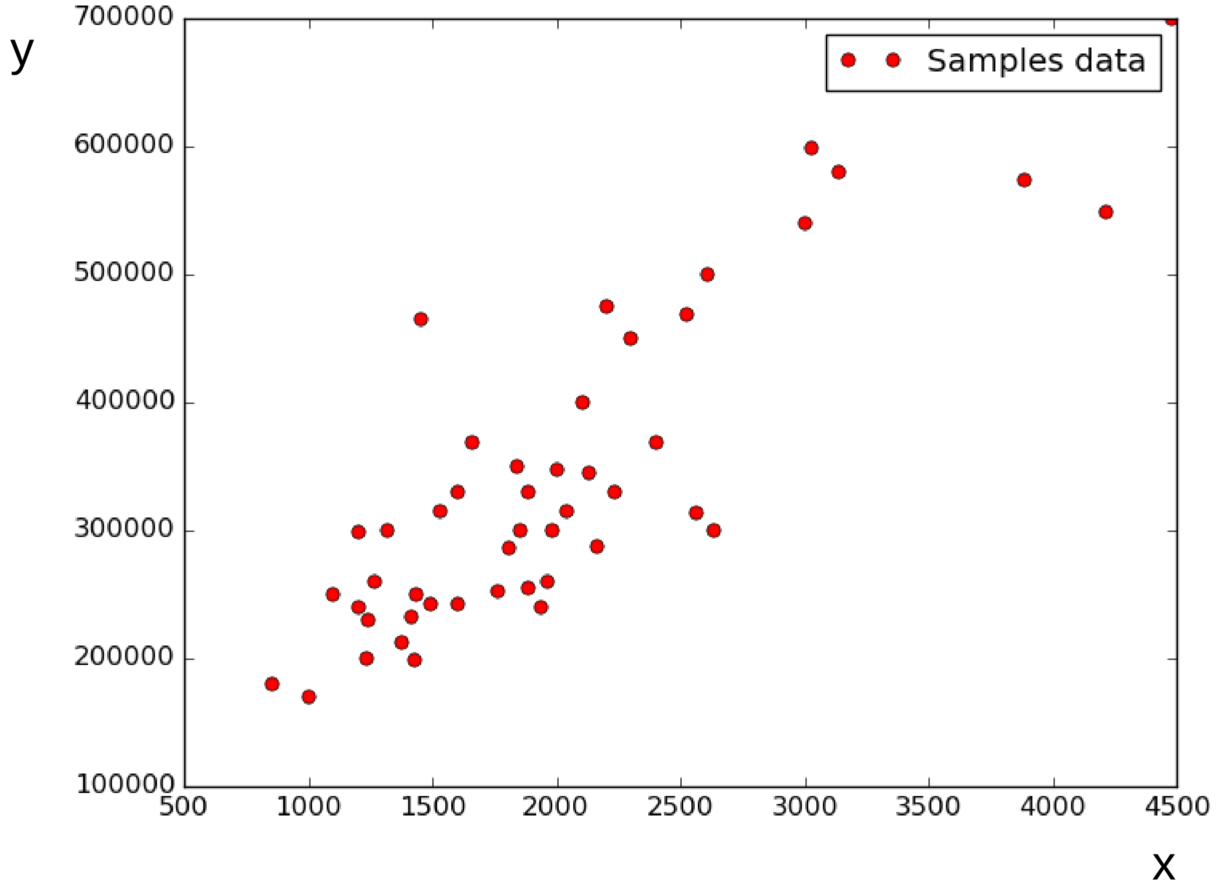
Regressão Linear

Método dos Mínimos
Quadrados

- Carl Friedrich Gauss (1809)
- Adrien M. Legendre (1805)

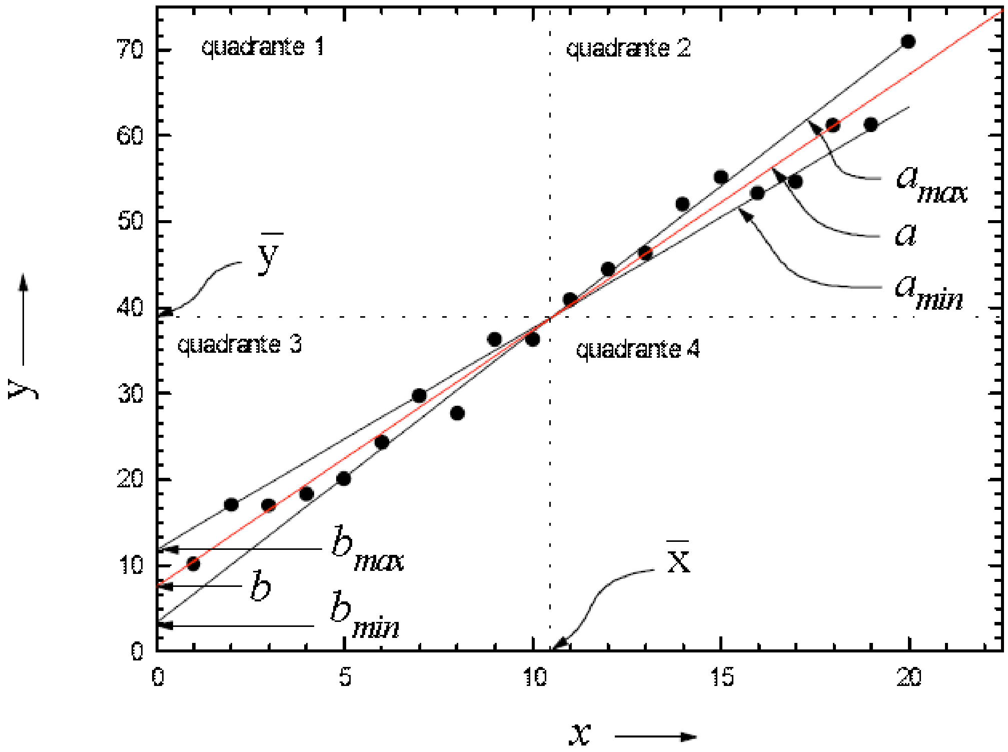
- Dados astronômicos e navegação:
- Como tratar medidas astronômicas obtidas nas mesmas condições ?

Situação real



https://www.altoros.com/blog/using-linear-regression-in-tensorflow/

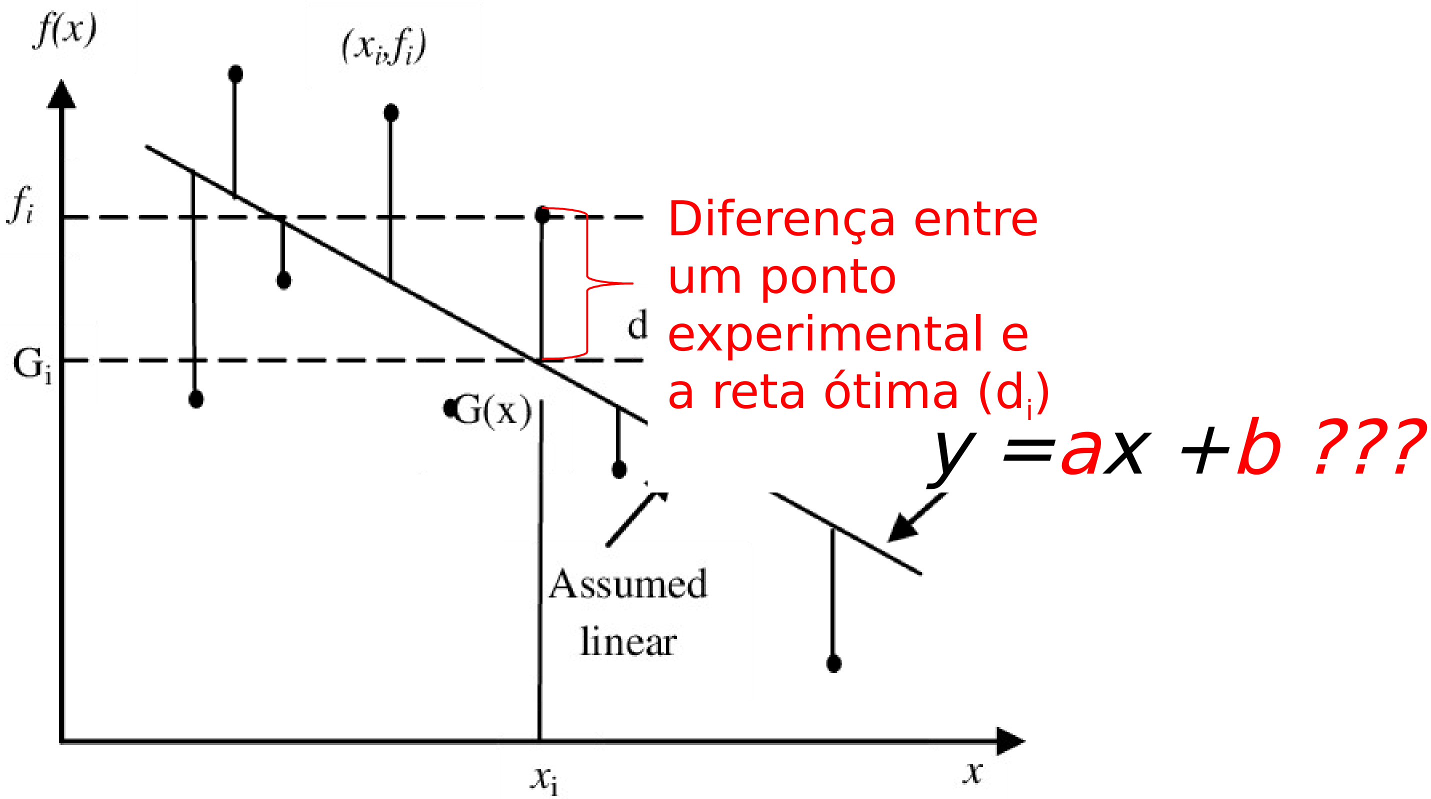
Qual é a melhor reta?



Fonte: Guia para Física Experimental
Caderno de Laboratório, Gráficos e Erros
Instituto de Física, Unicamp

Pergunta:

Qual o valor de **a** e **b** para se obter a melhor reta que descreve os dados?



https://www.researchgate.net/figure/Graphical-interpretation-of-the-least-squares-method_fig1_284367690

$$d_i = y_i - (ax_i + b)$$

Erros ou resíduos $d_i = y_i - (ax_i + b) = \quad e_i$

Vamos somar os erros: $\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - (ax_i + b)) = Soma_{\ erros}$

⇨ Problema!

Termos positivos e negativos se cancelam!

Soma Quadrática dos Erros

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 = SQ_{erros}$$

Como obter a melhor reta?

Minimizar SQ_{erros}

Variáveis: ***a*** e ***b***

$$\frac{\partial SQe}{\partial a} = 0$$

$$\frac{\partial SQe}{\partial b} = 0$$

Resolvendo essas derivadas, chegamos em:

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i^2) - n.\bar{x}^2}$$

$$\hat{b} = \bar{y} - \hat{a}.\bar{x}$$

- Encontramos **a** e **b**, então:
- **Y=ax+b**
- **Encontramos a reta ótima!**

- **Será necessário encontrar as incertezas de a e b.**

Além disso, temos:

Parâmetro	Estimador quando β = 0	Estimador quando β ≠ 0
α: coeficiente angular (inclinação)	$a = \Sigma x_i y_i / \Sigma x_i^2$	$a = \Sigma(x_i - \overline{x})y_i / \Sigma(x_i - \overline{x})^2$
β: coeficiente linear	$b = 0$	$b = \overline{y} - a\overline{x}$
σ^2: variância dos y_i	$S^2 = \Sigma(y_i - ax_i)^2 / (N-1)$	$S^2 = \Sigma(y_i - ax_i - b)^2 / (N-2)$
Δa: erro padrão do estimador da inclinação	$\Delta a = S / \sqrt{\Sigma x_i^2}$	$\Delta a = S / \sqrt{\Sigma(x_i - \overline{x})^2}$
Δb: erro padrão do estimador do coeficiente linear	$\Delta b = 0$	$\Delta b = S\sqrt{\frac{1}{N} + \frac{\overline{x}^2}{\Sigma(x_i - \overline{x})^2}}$
Δy_o: erro padrão do estimador de um ponto da reta para $x = x_o$	$\Delta y_o = Sx_o / \sqrt{\Sigma x_i^2}$	$\Delta y_o = S\sqrt{\frac{1}{N} + \frac{(x_o - \overline{x})^2}{\Sigma(x_i - \overline{x})^2}}$

Fonte: Guia para Física Experimental
Caderno de Laboratório, Gráficos e Erros
Instituto de Física, Unicamp

Coeficiente de correlação de Pearson

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt[2]{\sum_{i=1}^{n}(x_i - \bar{x})^2 . \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$\rho > 0.9$ Correlação linear forte entre x e y

$\rho < -0.9$ Anticorrelação linear forte entre x e y

$\rho \sim 0$ Baixa correlação entre x e y

Sugestão:

i	xi	yi	$x_i - \bar{x}$	x_i^2	$(x_i - \bar{x})y_i$	$(x_i - \bar{x})^2$
1	X1=1	Y1=3	1 -5			
2	X2=2	Y3=6	2-5			
...	...	...				
$\sum$						
médias	$\bar{x} = 5$					

Encontre a melhor reta que ajusta os dados abaixo:

x	y
1	5
2,0	7,5
3,0	9,0
4,5	12,0
5,0	13,0

Resposta:

Y=1,95x+3,25

Parâmetro	Estimador quando β = 0	Estimador quando β ≠ 0
α: coeficiente angular (inclinação)	$a = \Sigma x_i y_i / \Sigma x_i^2$	$a = \Sigma(x_i - \bar{x}) y_i / \Sigma(x_i - \bar{x})^2$
β: coeficiente linear	$b = 0$	$b = \bar{y} - a\bar{x}$
σ²: variância dos y_i	$S^2 = \Sigma(y_i - ax_i)^2 / (N-1)$	$S^2 = \Sigma(y_i - ax_i - b)^2 / (N-2)$
Δa: erro padrão do estimador da inclinação	$\Delta a = S / \sqrt{\Sigma x_i^2}$	$\Delta a = S / \sqrt{\Sigma(x_i - \bar{x})^2}$
Δb: erro padrão do estimador do	$\Delta b = 0$	[illegible]

- Dúvidas?