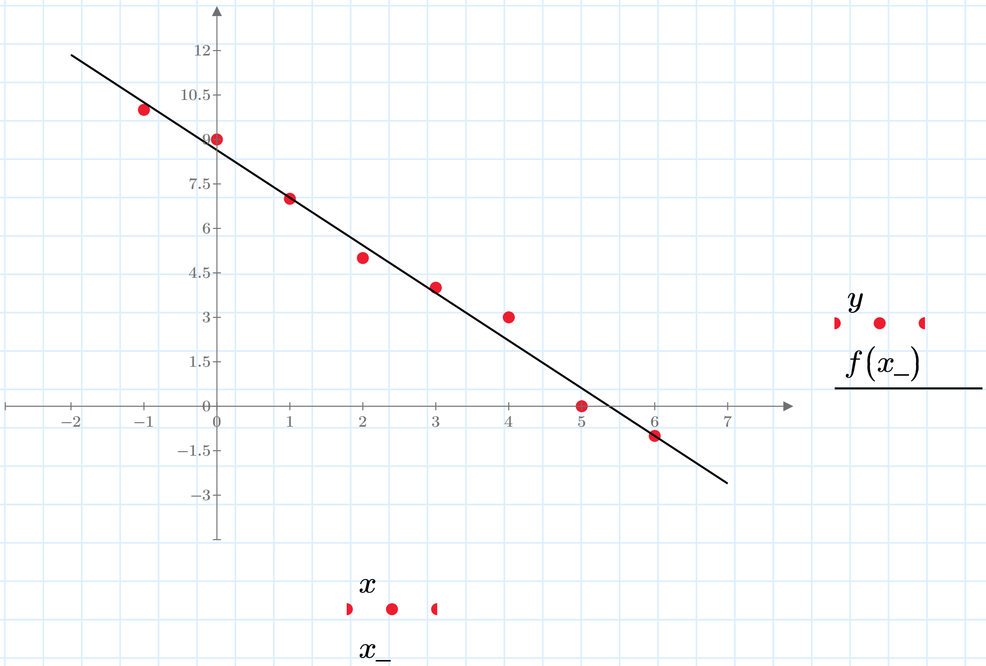
Least square curve fit

$$x := \begin{bmatrix} -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}^{T} \qquad y := \begin{bmatrix} 10 & 9 & 7 & 5 & 4 & 3 & 0 & -1 \end{bmatrix}^{T} \qquad n := \operatorname{length}(x) = 8$$

$$a_1 := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} \left(x_i\right)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = -1.607$$

$$a_0 := \operatorname{mean}(y) - a_1 \cdot \operatorname{mean}(x) = 8.643 \qquad f(x\_) := a_1 \cdot x\_ + a_0$$



$$r := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} \left(x_i\right)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n \sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} = -0.994$$

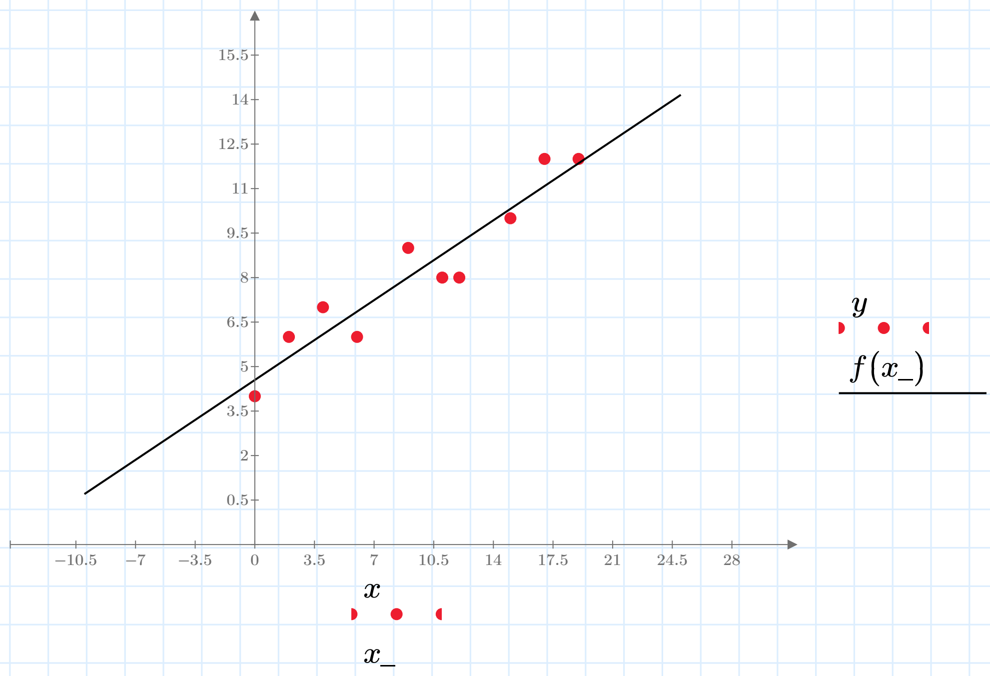
where **r** is the correlation, a measure of how the data points are correlated (follow a trend)

$$x := \begin{bmatrix} 0 & 2 & 4 & 6 & 9 & 11 & 12 & 15 & 17 & 19 \end{bmatrix}^{\mathrm{T}} \qquad y := \begin{bmatrix} 4 & 6 & 7 & 6 & 9 & 8 & 8 & 10 & 12 & 12 \end{bmatrix}^{\mathrm{T}}$$

$$n := \operatorname{length}(x) = 10$$

$$a_1 := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} \left(x_i\right)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = 0.385$$

$$a_0 := \operatorname{mean}(y) - a_1 \cdot \operatorname{mean}(x) = 4.547 \qquad f(x\_) := a_1 \cdot x\_ + a_0$$



$$r := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} \left(x_i\right)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n \sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} = 0.948$$

$$RSME := \sqrt{\frac{\sum_{i=1}^{n} \left(f\left(x_i\right) - y_i\right)^2}{n}} = 0.789$$

Where RSME is the root squared mean error

$$\textbf{clear}\left(a_1, a_0, r\right)$$

Linearization of non-linear r/ships

1. Power r/ship

$y = a \cdot x^{b}$ taking the logarithm from both sides, we get...

$log(y) = b \cdot log(x) + log(a)$

Now we treat log(y) as the dependent variable and log(x) as the independent variable with the free variables being b and log(a)

$x := \begin{bmatrix} 2.5 & 3.5 & 5 & 6 & 7.5 & 10 & 12.5 & 15 & 17.5 & 20 \end{bmatrix}^{T}$

$y := \begin{bmatrix} 13 & 11 & 8.5 & 8.2 & 7 & 6.2 & 5.2 & 4.8 & 4.6 & 4.3 \end{bmatrix}^{T}$ $\quad n := \text{length}(y) = 10$

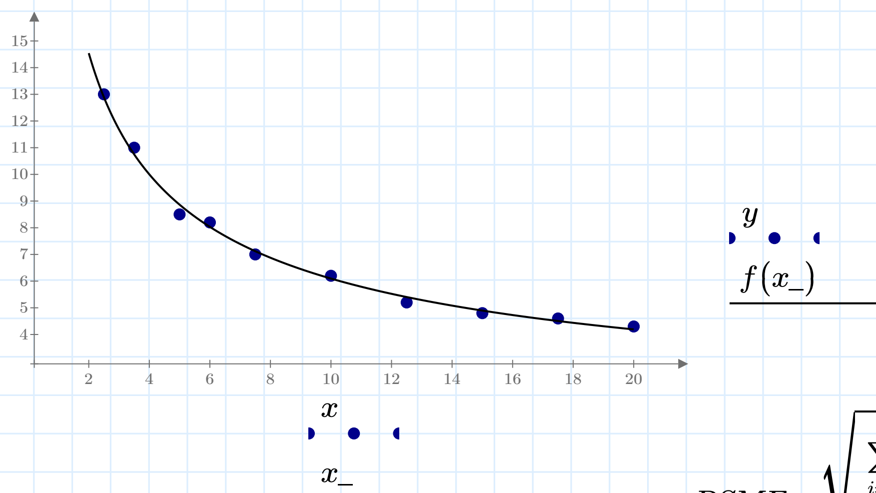
$y := \log(y) \quad x := \log(x)$ (Converting y and x to log_scale)

$$b := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} \left(x_i\right)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} {x_i}^{2} - \left(\sum_{i=1}^{n} x_i\right)^{2}} = -0.54$$

$$a' := \text{mean}(y) - b \cdot \text{mean}(x) = 1.325$$

$$a := a' = \log(a) \xrightarrow[]{solve,a \;\; float,4} 21.15$$

$f(x\_) := a \cdot x\_^{\,b} \quad y := 10^{y} \quad x := 10^{x}$ (Converting y and x back to true scale.)



$$RSME := \sqrt{\frac{\sum_{i=1}^{n} \left(f\left(x_i\right) - y_i\right)^{2}}{n}} = 0.182$$

clear (a, b)

2. Exponential r/ship

$y = a \cdot e^{b \cdot x}$ taking the natural logarithm from both sides, we get...

$ln(y) = b \cdot x + ln(a)$ Now we treat ln(y) as the independent variable without altering x.

$x := \begin{bmatrix} 0.4 & 0.8 & 1.2 & 1.6 & 2 & 2.3 \end{bmatrix}^{T}$ $y := \begin{bmatrix} 800 & 980 & 1500 & 1945 & 2900 & 3600 \end{bmatrix}^{T}$

$y := \ln(y)$ (Converting y to ln_scale) $n := \text{length}(x) = 6$

$$b := \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} (x_i)\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = 0.817$$

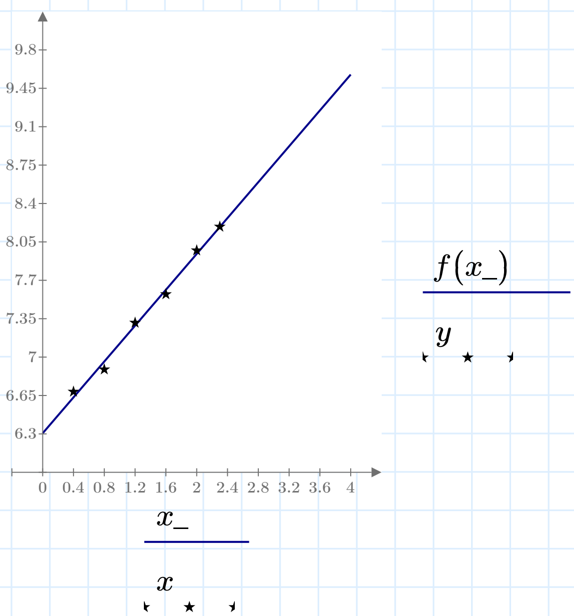
$a' := \text{mean}(y) - b \cdot \text{mean}(x) = 6.306$ $a := a' = \ln(a) \xrightarrow[\;]{solve, a \\ float, 3} 548.0$

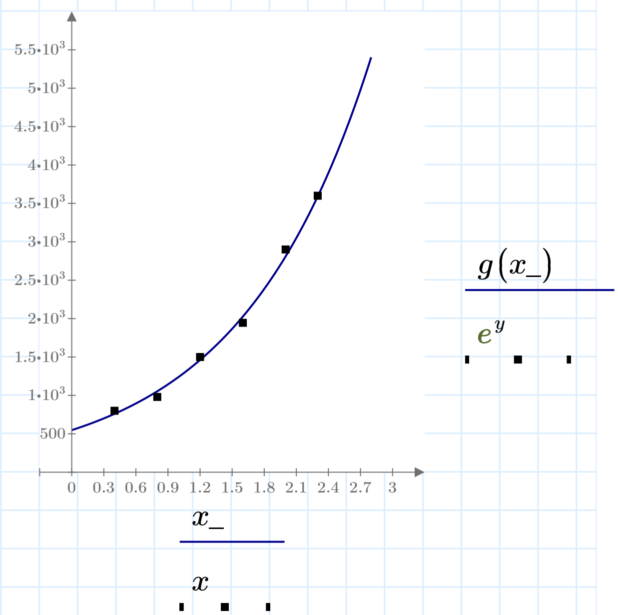
Model in the linear scale...

$f(x\_) := b \cdot x\_ + a'$

Model in the true (exponential) scale...

$g(x\_) := a \cdot e^{b \cdot x\_}$





$$RSME := \sqrt{\frac{\sum_{i=1}^{n} \left(f(x_i) - y_i\right)^2}{n}} = 0.043$$

Polynomial Curve fit

$x := \begin{bmatrix} -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}^{\mathrm{T}}$ $\quad$ $y := \begin{bmatrix} 10 & 6 & 2 & 1 & 0 & 2 & 4 & 7 \end{bmatrix}^{\mathrm{T}}$ $\quad$ $n := \mathrm{length}(x) = 8$

The formula for polynomial curve fit of any degree is...

$A \cdot a = b$ where **a** is a vector of coefficients and **A** and **b** are given as:

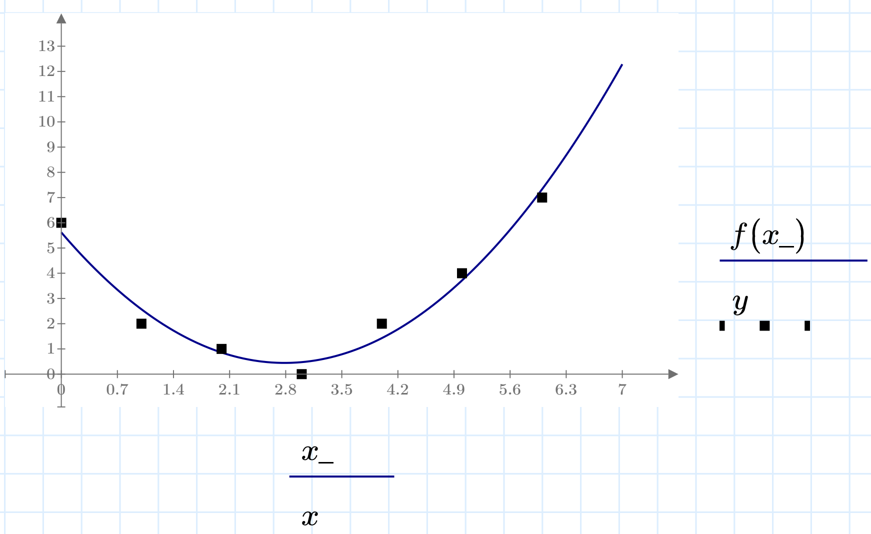
$$\mathbf{A} = \begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^{m-1} & \sum x_i^m & \sum x_i^{m+1} & \dots & \sum x_i^{2m} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^m y_i \end{bmatrix}$$

In our case, it's a quadratic curve fit so...

*__m__ is the degree of the polynomial

$$A := \begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} {x_i}^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} {x_i}^2 & \sum_{i=1}^{n} {x_i}^3 \\ \sum_{i=1}^{n} {x_i}^2 & \sum_{i=1}^{n} {x_i}^3 & \sum_{i=1}^{n} {x_i}^4 \end{bmatrix} \rightarrow \begin{bmatrix} 8 & 20 & 92 \\ 20 & 92 & 440 \\ 92 & 440 & 2276 \end{bmatrix} \qquad b := \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} \left(x_i \cdot y_i\right) \\ \sum_{i=1}^{n} \left({x_i}^2 \cdot y_i\right) \end{bmatrix} \rightarrow \begin{bmatrix} 32 \\ 64 \\ 400 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} := A^{-1} \cdot b \rightarrow \begin{bmatrix} \frac{118}{21} \\ -\frac{26}{7} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 5.619 \\ -3.714 \\ 0.667 \end{bmatrix} \qquad f(x\_) := a_0 + a_1 \cdot x\_ + a_2 \cdot {x\_}^2$$



Multiple Linear regression.

$x_1 := [0\ 0\ 1\ 2\ 0\ 1\ 2\ 2\ 1]^T$ $x_2 := [0\ 2\ 2\ 4\ 4\ 6\ 6\ 2\ 1]^T$ $n := \text{length}(x_1) = 9$

$y := [14\ 21\ 11\ 12\ 23\ 23\ 14\ 6\ 11]^T$

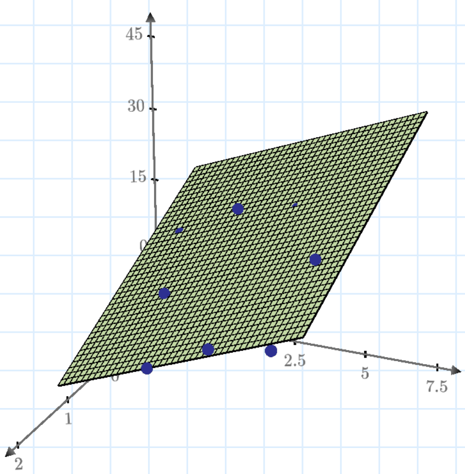
$$A := \begin{bmatrix} n & \sum_{i=1}^{n} x_{1_i} & \sum_{i=1}^{n} x_{2_i} \\ \sum_{i=1}^{n} x_{1_i} & \sum_{i=1}^{n} {x_{1_i}}^2 & \sum_{i=1}^{n} x_{1_i} \cdot x_{2_i} \\ \sum_{i=1}^{n} x_{2_i} & \sum_{i=1}^{n} x_{1_i} \cdot x_{2_i} & \sum_{i=1}^{n} {x_{2_i}}^2 \end{bmatrix} \rightarrow \begin{bmatrix} 9 & 9 & 27 \\ 9 & 15 & 33 \\ 27 & 33 & 117 \end{bmatrix}$$

$$b := \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1_i} \cdot y_i \\ \sum_{i=1}^{n} x_{2_i} \cdot y_i \end{bmatrix} \rightarrow \begin{bmatrix} 135 \\ 109 \\ 449 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} := A^{-1} \cdot b \rightarrow \begin{bmatrix} \frac{44}{3} \\ -\frac{20}{3} \\ \frac{7}{3} \end{bmatrix} = \begin{bmatrix} 14.667 \\ -6.667 \\ 2.333 \end{bmatrix}$$

$$f(x_{1\_}, x_{2\_}) := a_0 + a_1 \cdot x_{1\_} + a_2 \cdot x_{2\_}$$

$$Data := \text{augment}(x_1, x_2, y) = \begin{bmatrix} 0 & 0 & 14 \\ 0 & 2 & 21 \\ 1 & 2 & 11 \\ 2 & 4 & 12 \\ 0 & 4 & 23 \\ 1 & 6 & 23 \\ 2 & 6 & 14 \\ 2 & 2 & 6 \\ 1 & 1 & 11 \end{bmatrix}$$



$Data$

f

Now, our model is going to be a plane that best fits the 3D scatter plot of the Data.