Lecture 4. Least-squares fitting of a line and computation of correlation coefficient.
Sample data set #3, see lecture notes:

$$Y := \begin{pmatrix} 8.84 \\ 12.74 \\ 8.15 \\ 7.81 \\ 7.46 \\ 7.11 \\ 6.77 \\ 6.42 \\ 6.08 \\ 5.73 \\ 5.39 \end{pmatrix} \qquad X := \begin{pmatrix} 14 & 1 \\ 13 & 1 \\ 12 & 1 \\ 11 & 1 \\ 10 & 1 \\ 9 & 1 \\ 8 & 1 \\ 7 & 1 \\ 6 & 1 \\ 5 & 1 \\ 4 & 1 \end{pmatrix}$$

Data for a linear model

$N := 10$

$i := 0, 1 .. N$

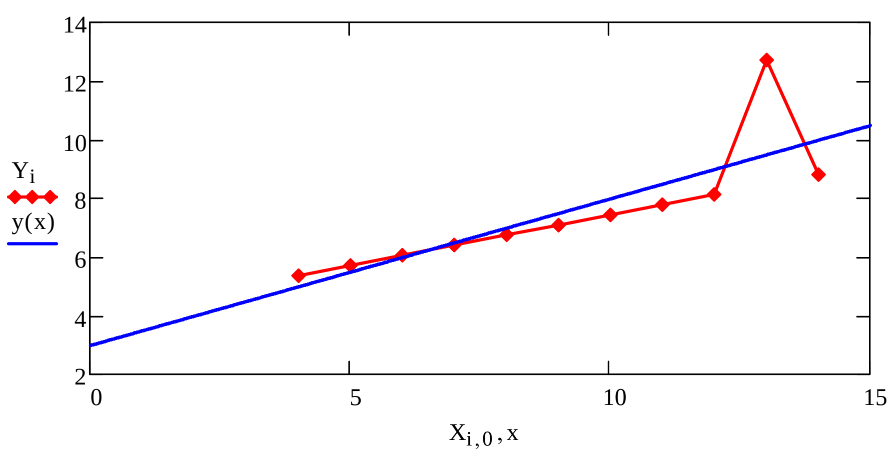
(note that there are 11 points)

$R := X^T \cdot Y \qquad M := X^T \cdot X$

$$R = \begin{pmatrix} 797.47 \\ 82.5 \end{pmatrix} \qquad M = \begin{pmatrix} 1.001 \times 10^3 & 99 \\ 99 & 11 \end{pmatrix} \qquad M^{-1} = \begin{pmatrix} 9.091 \times 10^{-3} & -0.082 \\ -0.082 & 0.827 \end{pmatrix}$$

$u := M^{-1} \cdot R \qquad u = \begin{pmatrix} 0.5 \\ 3.002 \end{pmatrix}$ (Vector of unknowns, i.e., slope and dc offset)

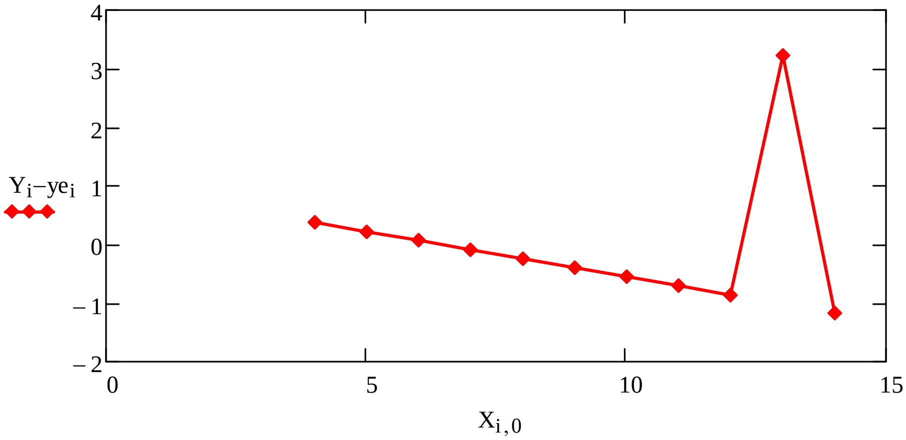
Fitted line: $y(x) := u_0 \cdot x + u_1$



Evaluate function: $ye_i := u_0 \cdot X_{i,0} + u_1$

(to compute values predicted by this linear model)

Plot of residuals (measured minus predicted):



Sum of the squares of residuals: $SSerr := \sum_{i=0}^{N} \left(Y_i - ye_i\right)^2$ $SSerr = 13.756$

$$Ymean := \frac{1}{N+1} \cdot \sum_{i=0}^{N} Y_i$$

Total sum of squares: $SStot := \sum_{i=0}^{N} \left(Y_i - Ymean\right)^2$ $SStot = 41.226$

Correlation coefficient (R2): $R2 := 1 - \frac{SSerr}{SStot}$ $R2 = 0.666$

Note that:

1) Second data point is not characteristic and should be discarded (use Chauvenet's criterion);

2) Correlation coefficient is a measure of how good mathematical model used (in this case a line) represents the measured data.