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

$$Y := \begin{pmatrix} 8.84 \\ 12.74 \\ 8.15 \\ 7.81 \\ 7.46 \\ 7.11 \\ 6.77 \\ 6.42 \\ 6.08 \\ 5.73 \\ 5.39 \end{pmatrix}$$

Data for a linear model

$$X := \begin{pmatrix} 14 & 1 \\ 13 & 1 \\ 12 & 1 \\ 11 & 1 \\ 10 & 1 \\ 9 & 1 \\ 8 & 1 \\ 7 & 1 \\ 6 & 1 \\ 5 & 1 \\ 4 & 1 \end{pmatrix}$$

$N := 10$

$i := 0, 1 .. N$

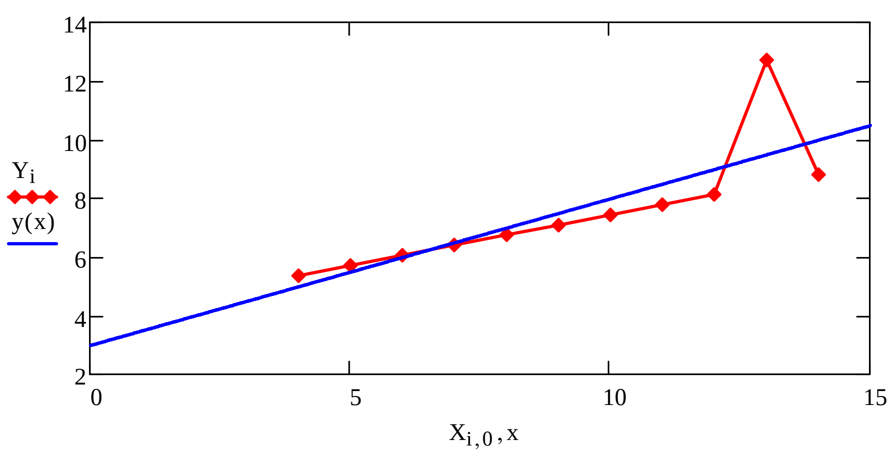
(note that there are 11 points)

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

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

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

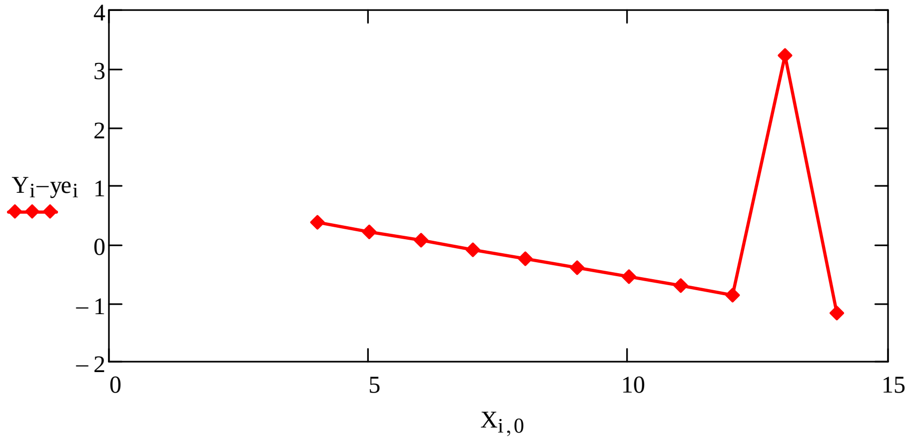
Fitted line: $\qquad y(x) := u_0 \cdot x + u_1$



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

(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.

Determination of +/- ranges for *slope* and *dc* offset: probability approach

Estimate standard deviation:

$Nparameters := 2$ (two parameters, slope and dc offset)

$DOF := (N + 1) - Nparameters$ $DOF = 9$ Degrees of freedom

$S := \sqrt{\frac{SSerr}{DOF}}$ $S = 1.236$

Use a 95% confidence level. Therefore,

$:= 0.05$

Consult *t-student* distribution:

$T \quad overTwo := 2.262$

$Minv := M^{-1}$

Ranges.

Upper range of slope:

$SlopeU := u_0 + T \quad overTwo \cdot S \cdot \sqrt{Minv_{0,0}}$ $SlopeU = 0.766$

Lower range of slope:

$SlopeL := u_0 - T \quad overTwo \cdot S \cdot \sqrt{Minv_{0,0}}$ $SlopeL = 0.233$

Upper range of dc offset:

$dcU := u_1 + T \quad overTwo \cdot S \cdot \sqrt{Minv_{1,1}}$ $dcU = 5.546$

Lower range of dc offset:

$dcL := u_1 - T \quad overTwo \cdot S \cdot \sqrt{Minv_{1,1}}$ $dcL = 0.459$

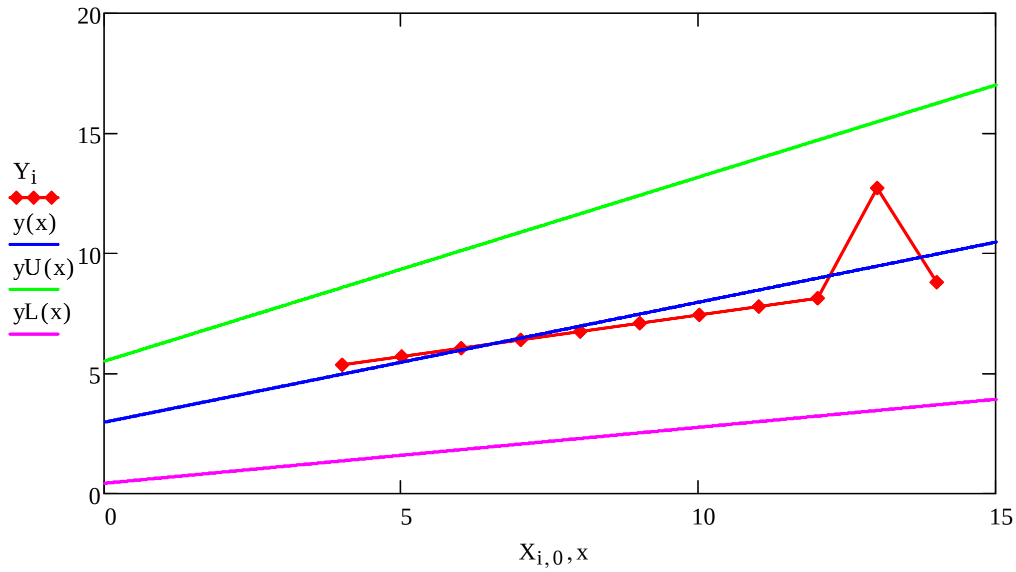
Equation of line, upper range:

$yU(x) := SlopeU \cdot x + dcU$

Equation of line, lower range:

$yL(x) := SlopeL \cdot x + dcL$

Plot of fitted line with +/- ranges (probability approach):



Note that *not* characteristic point has produced a very large range for both slope and dc values. Therefore, it is important to discard such a point (use Chauvenet's criterion)**.**