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

Data for a linear model

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

\[
N := 10 \\
i := 0, 1, \ldots, N
\]

(note that there are 11 points)

\[
R := X^T \cdot Y \\
M := X^T \cdot X
\]

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

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

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

Evaluate function: \(y(x) := u_0 \cdot X_{i,0} + u_1\)

(to compute values predicted by this linear model)
Plot of residuals (measured minus predicted):

Sum of the squares of residuals: \( SS_{err} := \sum_{i=0}^{N} (Y_i - ye_i)^2 \) \( SS_{err} = 13.756 \)

\[
Y_{\text{mean}} := \frac{1}{N+1} \sum_{i=0}^{N} Y_i
\]

Total sum of squares: \( SS_{tot} := \sum_{i=0}^{N} (Y_i - Y_{\text{mean}})^2 \) \( SS_{tot} = 41.226 \)

Correlation coefficient (R2): \( R^2 := 1 - \frac{SS_{err}}{SS_{tot}} \) \( R^2 = 0.666 \)

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:

\[ \text{Nparameters} := 2 \]  
\[ \text{(two parameters, slope and dc offset)} \]

\[ \text{DOF} := (N + 1) - \text{Nparameters} \]
\[ \text{DOF} = 9 \]  
\[ \text{Degrees of freedom} \]

\[ S := \sqrt{\frac{\text{SSerr}}{\text{DOF}}} \]
\[ S = 1.236 \]

Use a 95% confidence level. Therefore,

\[ := 0.05 \]

Consult \textit{t}-student distribution:

\[ T \ \text{overTwo} := 2.262 \]

\[ \text{M}^{-1} \]

Ranges.

Upper range of slope:

\[ \text{SlopeU} := u_0 + T \ \text{overTwo} \cdot S \cdot \sqrt{\text{M}^{-1}}_{0,0} \]
\[ \text{SlopeU} = 0.766 \]

Lower range of slope:

\[ \text{SlopeL} := u_0 - T \ \text{overTwo} \cdot S \cdot \sqrt{\text{M}^{-1}}_{0,0} \]
\[ \text{SlopeL} = 0.233 \]

Upper range of dc offset:

\[ \text{dcU} := u_1 + T \ \text{overTwo} \cdot S \cdot \sqrt{\text{M}^{-1}}_{1,1} \]
\[ \text{dcU} = 5.546 \]

Lower range of dc offset:

\[ \text{dcL} := u_1 - T \ \text{overTwo} \cdot S \cdot \sqrt{\text{M}^{-1}}_{1,1} \]
\[ \text{dcL} = 0.459 \]

Equation of line, upper range:

\[ y_U(x) := \text{SlopeU} \cdot x + \text{dcU} \]

Equation of line, lower range:

\[ y_L(x) := \text{SlopeL} \cdot x + \text{dcL} \]
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).