Given data: \((x_1, y_1), \ldots, (x_n, y_n)\)

Regressed data: \(\tilde{y}_i = mx_i + b\)

Residual error: \(r_i = y_i - \tilde{y}_i\)

Least squares problem:

Find \(m\) and \(b\) that would minimize the sum of residual error squared.

Let \(Q = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n} (y_i - mx_i - b)^2\)

At the minimum condition, we need the partial derivatives set to be zero:

\[
\frac{\partial Q}{\partial m} = 0 \quad \rightarrow \quad -2 \sum (y_i - mx_i - b)x_i = -2 (\sum y_i x_i - m \sum x_i^2 - b \sum x_i) = 0
\]

\[
\frac{\partial Q}{\partial b} = 0 \quad \rightarrow \quad -2 \sum (y_i - mx_i - b) = -2 (\sum y_i - m \sum x_i - n b) = 0
\]

Then,

\[
m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}
\]

\[
b = \frac{\sum y_i - m \sum x_i}{n}
\]

Coefficient of Determination: \(R^2\)

Let \(A = \sum_{i=1}^{n} (y_i - \bar{y})^2\) and \(B = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2\)

\[
R^2 = \frac{A - B}{A}
\]

Note: the smaller the value of \(B\), the closer \(R^2\) is to the value 1.
Standard Deviation of Residuals

\[ \sigma_r = \sqrt{\frac{\sum_{i=1}^{n} r_i^2}{n - 2}} \]

Standard Errors

\[ \Delta m = \sigma_r \sqrt{\frac{n}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}} \]

\[ \Delta b = \sigma_r \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}} \]

Confidence Intervals

Let \( k = t_{(1-\alpha),\nu} \) be the student t-score corresponding to the desired \((1 - \alpha)\) confidence level with degree of freedom, \(\nu = n - 2\). Then the \([(1 - \alpha) \times 100\%]\) confidence intervals for \(m\) and \(b\) are given by

\[ m \pm k \Delta m \quad \text{and} \quad b \pm k \Delta b \]