Suppose we have a data set $\{(x_i, y_i)\}_{i=1}^{N}$, where $x_j, y_j \in \mathbb{R}$. We derive the formula for the best linear fit $y = Ax + B$ for this dataset.

We assume that there is noise in our y-values, such that the $\{y_i\}_i$ are normally distributed around the "true" values, $A + Bx_i$, and that the normal distribution has the same variance σ^2 for each datapoint:

$$y_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - A - Bx_i)^2/2\sigma^2}.$$

We now wish to choose A and B to maximize the probability that we observe $\{y_i\}_i$ given $\{x_i\}_i$.

The probability of observing $\{y_i\}_i$ given $\{x_i\}_i$, A, and B is given by

$$p(y_1, \dots, y_N \mid x_1, \dots, x_N, A, B) =$$

$$p(y_1 \mid x_1, A, B) \cdot \dots \cdot p(y_N \mid x_N, A, B) =$$

$$\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2} \sum_i (y_i - A - Bx_i)^2}.$$

We now maximize this quantity with respect to A, B.

$$\frac{\partial p}{\partial A} = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_i (y_i - A - Bx_i)^2} \cdot \left(-\frac{1}{2\sigma^2}\right).$$

$$\sum_i 2(y_i - A - Bx_i)\cdot(-1) = 0 \iff$$

$$NA + B\sum_i x_i = \sum_i y_i .$$

$$\frac{\partial p}{\partial B} = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_i (y_i - A - Bx_i)^2} \cdot \left(-\frac{1}{2\sigma^2}\right).$$

$$\sum_i 2(y_i - A - Bx_i)(-x_i) = 0 \iff$$

$$\sum_i Ax_i + Bx_i^2 = \sum_i x_i y_i .$$

These equations can now be rewritten

$$\begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} \iff$$

$$A = \frac{1}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\left[\begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & N \end{pmatrix}\begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}\right]_1$$

$$\rightarrow \boxed{A = \frac{1}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\left[\sum_i x_i^2 \sum_j y_j - \sum_i x_i \sum_j x_j y_j\right]}$$

$$B = \frac{1}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\left[\begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & N \end{pmatrix}\begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}\right]_2$$

$$\rightarrow \boxed{B = \frac{1}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}\left[-\sum_i x_i \sum_j y_j + N\sum_i x_i y_i\right]}$$

(Source: Taylor, Error Analysis, p. 182-184)