We derive the formula for linear regression in one independent and one dependent variable.

(Source: Numerical Recipes, and Taylor Error Analysis)

Weighted Linear Regression

Suppose we have N data points $(x_i, y_i)_{i=1}^{N}$, and we believe that they are linearly related by

$$y(x) = a + bx.$$

We show how to find the values of a and b that best fit the dataset. **?**

In our simulations, we also know the uncertainties σ_i associated with each y_i, and we will assume that the uncertainty in the x_i's are negligible. We can then introduce chi-squared function for our dataset

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - a - bx_i}{\sigma_i} \right)^2 .$$

The optimal parameter values a and b will minimize this function. Equating derivatives to zero, we have

$$\frac{\partial \chi^2}{\partial a} = 0 = \sum_{i=1}^{N} \frac{y_i - a - bx_i}{\sigma_i^2}$$

$$\frac{\partial \chi^2}{\partial b} = 0 = \sum_{i=1}^{N} \frac{x_i(y_i - a - bx_i)}{\sigma_i^2}.$$

Next, we introduce the shorthand

$$S \equiv \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

$$S_x \equiv \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}$$

$$S_y \equiv \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}$$

$$S_{xx} \equiv \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2}$$

$$S_{xy} \equiv \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2}$$

which gives us equations for a and b: $aS + bS_x = S_y$ and $aS_x + bS_{xx} = S_{xy}$, or equivalently

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{SS_{xx} - S_x^2} \begin{pmatrix} S & S_x \\ S_x & S_{xx} \end{pmatrix}^{-1} \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \equiv \frac{1}{\Delta} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix} \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \implies$$

$$a = (S_{xx}S_y - S_xS_{xy})/\Delta \tag{1}$$
$$b = (SS_{xy} - S_xS_y)/\Delta. \tag{2}$$

These are the best-fit parameter values. We now estimate their uncertainty. We may write the best-fit parameters as

$$a = a(x_1, ..., x_N, y_1, ..., y_N)$$
$$b = b(x_1, ..., x_N, y_1, ..., y_N),$$

where we recall that the uncertainty in x_i is negligible and the uncertainty in y_i is σ_i. Then, the standard formulas of error propagation **?** give an uncertainty of

$$\sigma_a^2 = \sum_{i=1}^{N} \left(\frac{\partial a}{\partial y_i}\right)^2 \sigma_i^2 = S_{xx}/\Delta$$
$$\sigma_b^2 = \sum_{i=1}^{N} \left(\frac{\partial b}{\partial y_i}\right)^2 \sigma_i^2 = S/\Delta.$$

Weighted Linear Regression with No Intercept

Suppose we have N data points $(x_i, y_i)_{i=1}^N$, and we believe that they are linearly related by

$$y(x) = bx.$$

We show how to find the values of a and b that best fit the dataset.

In our simulations, we also know the uncertainties σ_i associated with each y_i, and we will assume that the uncertainty in the x_i's are negligible. We can then introduce chi-squared function for our dataset

$$\chi^2(b) = \sum_{i=1}^{N} \left(\frac{y_i - bx_i}{\sigma_i}\right)^2.$$

The optimal parameter value b will minimize this function. Equating the derivative to zero, we have

$$\frac{\partial \chi^2}{\partial b} = 0 = \sum_{i=1}^{N} \frac{x_i(y_i - bx_i)}{\sigma_i^2} \iff S_{xy} = bS_{xx} \iff b = S_{xy}/S_{xx}.$$

This is the best-fit parameter value for b. We now estimate its uncertainty. Using the same relation above, we have

$$\sigma_b^2 = \sum_{i=1}^{N} \left(\frac{\partial b}{\partial y_i}\right)^2 \sigma_i^2 = \sqrt{1/S_{xx}}.$$