Simple program for fitting a straight line to data

We often encounter the problem of finding the line that best reproduces (fits) the behavior of numerical data (from calculations or experiments). Often the data have error bars, i.e., we have points x(i),y(i),s(i), where i=1,....,N and s(i) is the standard deviation of y(i).

To find the **best-fit line** $y = a + bx$ we minimize χ^2

$$\chi^2 = \sum_{i=1}^{N} (y_i - a - bx_i)^2/\sigma_i^2$$

We thus have to find a and b such that

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 0 \\ \frac{\partial \chi^2}{\partial b} &= 0 \end{aligned}$$

This gives

$$\begin{aligned} a\Sigma(1) + b\Sigma(x) - \Sigma(y) &= 0 \\ a\Sigma(x) + b\Sigma(x^2) - \Sigma(xy) &= 0 \end{aligned}$$

where we have defined

$$\begin{aligned} \Sigma(1) &= \sum_i 1/\sigma_i^2 \\ \Sigma(x) &= \sum_i x_i/\sigma_i^2 \\ \Sigma(y) &= \sum_i y_i/\sigma_i^2 \\ \Sigma(x^2) &= \sum_i x_i^2/\sigma_i^2 \\ \Sigma(xy) &= \sum_i x_i y_i/\sigma_i^2 \end{aligned}$$

Solving for a and b gives

$$\begin{aligned} b &= \frac{\Sigma(xy)\Sigma(1) - \Sigma(x)\Sigma(y)}{\Sigma(x^2)\Sigma(1) - [\Sigma(x)]^2} \\ a &= \frac{\Sigma(y) - b\Sigma(x)}{\Sigma(1)} \end{aligned}$$

The **statistical errors** σ_i of the y-value y_i characterizes its fluctuations (or uncertainty). To calculate the statistical errors of the parameters a and b we have to calculate the shifts of these parameters as the y-values are fluctuating independently and sum the squares of these shifts (i.e., the variances) to obtain the variance of the parameters. The errors (standard deviations) are the square-roots of the variances. The statistical errors of the parameters are thus given by

$$\sigma_a = \sqrt{\sum_i (\delta a_i)^2}$$
$$\sigma_b = \sqrt{\sum_i (\delta b_i)^2}$$

where δa_i and δb_i are the leading-order shifts

$$\delta a_i = \sigma_i \frac{\partial a}{\partial y_i}$$
$$\delta b_i = \sigma_i \frac{\partial b}{\partial y_i}$$

The derivatives are

$$\frac{\partial b}{\partial y_i} = \frac{(x_i/\sigma_i^2)\Sigma(1) - (1/\sigma_i^2)\Sigma(x)}{\Sigma(x^2)\Sigma(1) - [\Sigma(x)]^2}$$
$$\frac{\partial a}{\partial y_i} = \frac{(1/\sigma_i^2) - (\partial b/\partial y_i)\Sigma(x)}{\Sigma(1)}$$

This is a Fortran 90 line-fitting program [linefit.f90]

Running the brogram with this data set gives the output

```
Intercept and expected error :  0.22112252494404125 0.009332726022970368
Slope and expected error     :  0.5506251596885045 0.002800789912660865
Number of data points        :  51
X2 per degree of freedom     :  1.170613622829978
```

The data was generated by adding noise (distributed according to Gaussians with mean 0 and variable width equal to the stated errors) to a line y=a+bx with a=0.23 and b=0.55. These parameters are seen to be reproduced by the fitting program to within statistical errors.

This graph of the data was generated using Xmgrace