1 Data Prep

First we need to convert distance modulus to distance. Note that

$$d(\mu) = 10^{1+\frac{\mu}{5}}$$

that means that
$$\frac{dd}{d\mu} = \left(\frac{\log(10)}{5}\right)\left(10^{1+\frac{\mu}{5}}\right)$$

or in other words the uncertainties need to be propagated

$$\sigma_d = \frac{dd}{d\mu}\sigma_\mu$$

Once we have processed the data we end up with a dataset with points

$$(x_i, y_i) \in S$$

2 Linear regression

For this data we want to fit a function

$$y = Ax + b$$

to extract the trend in this data. To do this we want this equation to predict the a position $\hat{y}_i$ for some point x_i that is closest to the actual true point. As a consequence, we want to minimize the equation

$$Q = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

where we have
$$\hat{y}_i = Ax_i + b$$

or in other words we want to find A and b such that Q is minimized, where now

$$Q = \sum_{i=1}^{N} (y_i - Ax_i - b)^2$$

To do this we just take a derivative for both A and b

$$\frac{dQ}{dA} = \sum_{i=1}^{N} -2x_i (y_i - Ax_i - b) = \sum_{i=1}^{N} -2\left(x_i y_i - Ax_i^2 - bx_i\right) = 0$$
$$\frac{dQ}{db} = \sum_{i=1}^{N} -2(y_i - Ax_i - b) = 2Nb + 2A\sum_{i=1}^{N} x_i - 2\sum_{i=1}^{N} y_i = 0$$

Now from $\frac{dQ}{db} = 0$ we can solve for b

$$\begin{aligned} b &= \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{A}{N}\sum_{i=1}^{N} x_i \\ &= \bar{y} - A\bar{x} \end{aligned}$$

where we have defied the average of all y_i as $\bar{y}$ and the average of all x_i as $\bar{x}$ Now lets substitute definition of b into $\frac{dQ}{dA}$ then we have

$$\begin{aligned} \frac{dQ}{dA} &= \sum_{i=1}^{N} -2\left(x_i y_i - A x_i^2 - (\bar{y} - A\bar{x})\, x_i\right) \\ &= \sum_{i=1}^{N} -2\left(x_i y_i - \bar{y} x_i + A\bar{x} x_i - A x_i^2\right) \\ &= -2\sum_{i=1}^{N} x_i\,(y_i - \bar{y}) - 2A\sum_{i=1}^{N} x_i\,(\bar{x} - x_i) \end{aligned}$$

and so solving for A we have

$$A = \frac{\sum_{i=1}^{N} x_i\,(y_i - \bar{y})}{\sum_{i=1}^{N} x_i\,(x_i - \bar{x})} = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}$$

which we can rewrite noting $\sum_{i=1}^{N}\left(\bar{x}^2 - x_i\bar{x}\right) = \sum_{i=1}^{N}\left(\bar{x}\bar{y} - y_i\bar{x}\right) = 0$

$$A = \frac{\sum_{i=1}^{N} x_i\,(y_i - \bar{y}) + \sum_{i=1}^{N}(\bar{x}\bar{y} - y_i\bar{x})}{\sum_{i=1}^{N} x_i\,(x_i - \bar{x}) + \sum_{i=1}^{N}(\bar{x}^2 - x_i\bar{x})} = \frac{\sum_{i=1}^{N} x_i y_i - x_i\bar{y} + \bar{x}\bar{y} - y_i\bar{x}}{\sum_{i=1}^{N} x_i x_i - \bar{x}x_i + \bar{x}^2 - x_i\bar{x}}$$

$$A = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\mathrm{COV}(x,y)}{\mathrm{VAR(x)}}$$

Where VAR is the variance, and the other variable COV is what we call the covariance, which is defined by

$$\mathrm{COV(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}$$

Given this parameter set we can also define the correlation coeficient r^2 which is very closely related to A. This we write as

$$r^2 = \frac{\mathrm{COV}^2(x,y)}{\mathrm{VAR(x)VAR(y)}}$$

and we have

$$r = A\sqrt{\frac{\mathrm{VAR(y)}}{\mathrm{VAR(x)}}}$$

Now what about the uncertainty on A and b. To do that lets first comput the RMS of our distribution, this we often refer to as the residual sum of the squares (RSS)

$$\mathrm{RSS} = \sum_{\mathrm{i=1}}^{\mathrm{N}} \left(\mathrm{y_i} - \mathrm{f(x_i)}\right)^2 = \sum_{\mathrm{i=1}}^{\mathrm{N}} \left(\mathrm{y_i} - \mathrm{Ax_i} + \mathrm{b}\right)^2$$

this we often write in terms of mean squared error (MSE)

$$MSE = \hat{\sigma}^2_{MSE} = \frac{1}{N-2} \sum_{i=1}^{N} \left(y_i - Ax_i - b\right)^2$$

Note that the $\frac{1}{N-2}$ is to account for the fact that A and b are determined from the data, and thus remove 2 degress of freedom. To understand this imagine what the MSE would be if you fit 2 points (0), so in fact there are no degrees of freedom of variance. A third point would thus fluctuate about the line with an MSE consistent of one point fluctions. Now we want to compute the variance of A. To do this lets go back to the definition of A

$$A = \frac{\frac{1}{N}\sum_{i=1}^{N} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\frac{1}{N}\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2} = \sum_{i=1}^{N} \frac{\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2} \left(y_i - \bar{y}\right) = \sum_{i=1}^{N} w_i \left(y_i - \bar{y}\right)$$

Consequently the variance of A is the sum of the variances of each of y_i

$$\mathrm{VAR(A)} = \sum_{\mathrm{i}} \mathrm{w_i^2} \left(y_i - \hat{y}_i\right)^2 = RSS \sum_i w_i^2 = \mathrm{VAR(y)} \sum_i \left(\frac{\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}\right)^2$$

$$\mathrm{VAR(A)} = \mathrm{VAR(y)} \frac{\sum_i \left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2\right)^2} = \mathrm{VAR(y)} \frac{1}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}$$

$$\mathrm{VAR(A)} = \mathrm{VAR(y)} \frac{1}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2} = \frac{\mathrm{VAR(y)}}{N\mathrm{VAR(x)}}$$

Now to correct for the fact that A and b are constrained from the data we have

$$\mathrm{VAR}(A) = \frac{\mathrm{VAR(y)}}{\left(N-2\right)\mathrm{VAR(x)}}$$

Additionally, the variance of b is given by considering the fact that

$$b = \bar{y} - A\bar{x}$$

thus we have that

$$\begin{aligned} \sigma_b^2 &= \sigma_{\bar{y}}^2 + \sigma_a^2 \bar{x}^2 \\ &= \frac{1}{N} \mathrm{VAR(y)} + \mathrm{VAR(A)} \bar{\mathrm{x}}^2 \end{aligned}$$

$$\mathrm{VAR(b)} = \frac{\hat{\sigma}^2_{\mathrm{MSE}}}{\mathrm{N}} + \mathrm{VAR(A)}\bar{\mathrm{x}}^2$$

Ok, this is the only time we are going to do things analytically.

3 Weighted Linear regression

With uncertainties on each point we know have

$$Q = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2}$$

and thus

$$Q = \sum_{i=1}^{N} \frac{(y_i - Ax_i - b)^2}{\sigma_i^2}$$

$$\frac{dQ}{dA} = \sum_{i=1}^{N} -2x_i \left(y_i - Ax_i - b\right) \frac{1}{\sigma_i^2} = \sum_{i=1}^{N} -2 \left(x_i y_i - Ax_i^2 - bx_i\right) \frac{1}{\sigma_i^2} = 0$$

$$\frac{dQ}{db} = \sum_{i=1}^{N} -2 \left(y_i - Ax_i - b\right) \frac{1}{\sigma_i^2} = 2Nb \sum_{i} \frac{1}{\sigma_i^2} + 2A \sum_{i=1}^{N} x_i \frac{1}{\sigma_i^2} - 2 \sum_{i=1}^{N} y_i \frac{1}{\sigma_i^2} = 0$$

We can solve for this skipping some math and defining the weighted means

$$\bar{y}_w = \frac{\sum_i \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

$$b = \bar{y}_w - A\bar{x}_w$$

$$A = \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(x_i - \bar{x}_w\right)\left(y_i - \bar{y}_w\right)}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(x_i - \bar{x}_w\right)^2}$$

$$\sigma_b^2 = \left(\frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} + \frac{\bar{x}_w^2}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(x_i - \bar{x}_w\right)^2} \right) \sigma^2$$

$$\sigma_A^2 = \frac{\sigma^2}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(x_i - \bar{x}_w\right)^2}$$

with

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

4 Quick Aside

While we did the whole thing above for just a linear function. You could easily do this for a vector, namely

$$\vec{y} = A\vec{x} + \vec{b}$$

where A is now a matrix. This is often rewritteh as

$$\vec{y} = \beta\vec{x}'$$

where $\vec{x}' = (\vec{x}, 1)$. Also, the variables $\vec{y}$ and $\vec{x}$ are recentered so that their mean is zero. The resulting minimum is the generalized version of above

$$\hat{\beta} = \left(\vec{x}^T \vec{x}\right)^{-1} \vec{x}^T \vec{y}$$

You can add weights too

5 Minimizing without all the math

What if we want to minimize with a generic function?

Lets start with minimizing the function below

$$Q = \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2$$

Now instead of assuming a linear solution we assume a function form

$$Q = \sum_{i=1}^{N} \left(y_i - f(x)\right)^2$$

This function will have a number of free parameters like with the linear fit we had A and b. The standard way to write this is $f(x|\theta_i)$ where θ_i is a collection of parameters that you wish to fit. To do this what we want to find is

$$\frac{\partial Q}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \sum_{i=1}^{N} \left(y_i - f(x|\theta_i)\right)^2 = \sum_{i=1}^{N} 2\left(y_i - f(x|\theta_i)\right) \frac{\partial f(x|\theta_i)}{\partial \theta_i} = 0$$

The trick is we need to this simultaneously for all θ_i. In the simple case