Lab 7 Addendum: Calculating regression using least squares method

If we want to solve a regression equation

$$y = a_0 + a_1 x + a_2 x_2 + \ldots$$

we can write it in matrix form as:

$$\begin{bmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,\ldots} \\ 1 & x_{2,1} & x_{2,2} & x_{2,\ldots} \\ 1 & x_{3,1} & x_{3,2} & x_{3,\ldots} \\ . & x_{\ldots,1} & x_{\ldots,2} & x_{\ldots,\ldots} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_{\ldots} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_{\ldots} \end{bmatrix}$$

where each column is all of the data points for that series, i.e. first a column of ones, then all of the x1 values etc as below:

$$\begin{bmatrix} 1 & x_1 values & x_2 values \end{bmatrix} \begin{bmatrix} a's \end{bmatrix} = \begin{bmatrix} yvalues \end{bmatrix}$$

This can also be written

$$[X][A] = [Y]$$

$$X * A = Y$$

(traditionally this matrix problem would be written as A*x=b but that is confusing so we will adjust our notation). We can solve this by:

$$A = inv(X^T X) * X^T * Y$$

where T means transpose of a matrix.

Excel can solve this using a matrix multiply function (mmult) the matrix inverse function (minverse) and transpose (transpose).

A=mmult(mmult(minverse(mmult(transpose(X),X),transpose(X)),Y)

You need to select an output range in a column that is as many rows as there are x variables and execute using command-enter (control-shift-enter on PC). The values in A represent the coefficients a0, a1 etc. and are equivalent to the coefficients we calculated after converting our equation from z units back to the original units.

This method is easier to use in my opinion and is versatile because you can do both simple and multiple regression and can add variables just by adding a column to X and a row to the output.