Linear Regression
(Least Squares Method)

We have briefly talked about the correlation coefficient, r^2. This r^2 is a measure of how well a set of data fits a linear fit. The value of r^2 ranges from 0 to 1, a value of 1 means that the data is dead on linear and that a value of 0 means that the data is not linear. (R-squared value: An indicator from 0 to 1 that reveals how closely the estimated values for the trend line correspond to your actual data. A trend line is most reliable when its R-squared value is at or near 1. Also known as the coefficient of determination. Defined by Microsoft Excel.) The r^2 value is determined by squaring the Pearson product moment correlation coefficient (PPMC) which is symbolized as r. Using equation 1, you can calculate the r. In order to do this it makes life easy if you do some calculations ahead of time.

Let us use an imagery set of data. It is always best that we organize the data in columns. The x axis data is normally positioned in the first column while the y axis data goes in the second column.

x	y
-3.8	8.2
-1.2	6
0.75	3.3
3.1	1.6
5.6	-2
8	-6

	x	y	xy	x^2	y^2
	-3.8	8.2	-31.16	14.44	67.24
	-1.2	6	-7.2	1.44	36
	0.75	3.3	2.475	0.5625	10.89
	3.1	1.6	4.96	9.61	2.56
	5.6	-2	-11.2	31.36	4
	8	-6	-48	64	36
Σ =	12.45	11.1	-90.13	121.41	156.69

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[n(\sum x^2) - (\sum x)^2\right]\left[n(\sum y^2) - (\sum y)^2\right]}} \quad (1)$$

$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \quad (2)$$

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

Where n is equal to the number of data points.
The goal of this lab is two fold.

1. Using your imaginary set of data and equations 1, 2, and 3, calculate the r, r^2, y-intercept (b), and slope (m) of your imaginary data. This can be done one of two ways
    a. Using a spreadsheet program, such as Microsoft Excel
    b. By hand
2. Graphing your imaginary set of data using
    a. Microsoft Excel
    b. By hand

What your data will look like in a spread sheet

	A	B	C	D	E
1	**x**	**y**	**xy**	**x^2**	**y^2**
2	-3.8	8.2	-31.16	14.44	67.24
3	-1.2	6	-7.2	1.44	36
4	0.75	3.3	2.475	0.5625	10.89
5	3.1	1.6	4.96	9.61	2.56
6	5.6	-2	-11.2	31.36	4
7	8	-6	-48	64	36
8	12.45	11.1	-90.125	121.4125	156.69
9					
10					
11					
12	**m(top)**	-678.945	-1.18392	**m**	
13	**m(bottom**	573.4725			
14					
15	**r(top)**	-678.945	-0.99194	**r**	
16	**r(bottom)**	684.4610211			
17					
18	**r^2**	0.983947093			
19					
20	**b(top)**	2469.735	4.306632	**b**	
21	**b(bottom)**	573.4725			

What to type in to manipulate the data. Make sure to use the power of a spreadsheet and your organization. I will show you briefly how to do this using cut & paste and some other tricks.

	A	B	C	D	E
1	**x**	**y**	**xy**	**x^2**	**y^2**
2	-3.8	8.2	=B2*A2	=A2^2	=B2^2
3	-1.2	6	=B3*A3	=A3^2	=B3^2
4	0.75	3.3	=B4*A4	=A4^2	=B4^2
5	3.1	1.6	=B5*A5	=A5^2	=B5^2
6	5.6	-2	=B6*A6	=A6^2	=B6^2
7	8	-6	=B7*A7	=A7^2	=B7^2
8	=SUM(A2:A7)	=SUM(B2:B7)	=SUM(C2:C7)	=SUM(D2:D7)	=SUM(E2:E7)
9					
10					
11					
12	**m(top)**	=6*C8-B8*A8	=B12/B13	**m**	
13	**m(bottom)**	=6*D8-A8^2			
14					
15	**r(top)**	=6*C8-A8*B8	=B15/B16	**r**	
16	**r(bottom)**	=SQRT((6*D8-A8^2)*(6*E8-B8^2))			
17					
18	**r^2**	=C15^2			
19					
20	**b(top)**	=B8*D8-A8*C8	=B20/B21	**b**	
21	**b(bottom)**	=6*D8-A8^2			

What the graph will look like.