

**Linear Regression with Excel** (explicitly for Excel 2010, but the process is very similar for other versions of Excel):

(taken from <http://blog.vojimbocorp.com/2012/05/03/linear-regression-with-excel-2010/>)

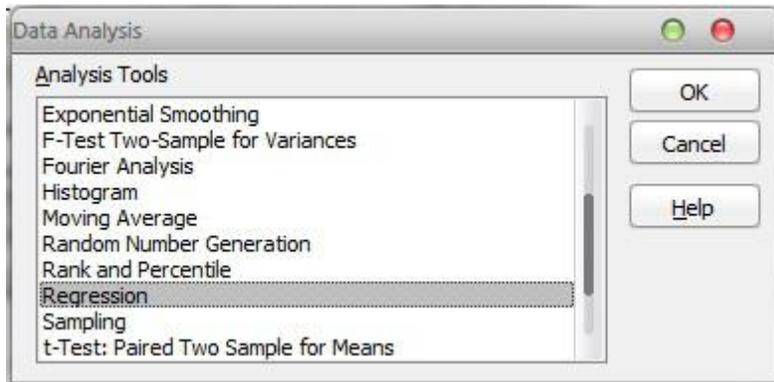
First, we will need to enable the Analysis ToolPak for Excel:

- 1) Click File and select options
- 2) Select Add-ins on the left hand menu
- 3) Select Excel Add-ins in the drop down list named Manage at the bottom of the pop up
- 4) Click the 'Go' button
- 5) Tick the checkbox for Analysis ToolPak if it is empty
- 6) Click the 'Ok' button

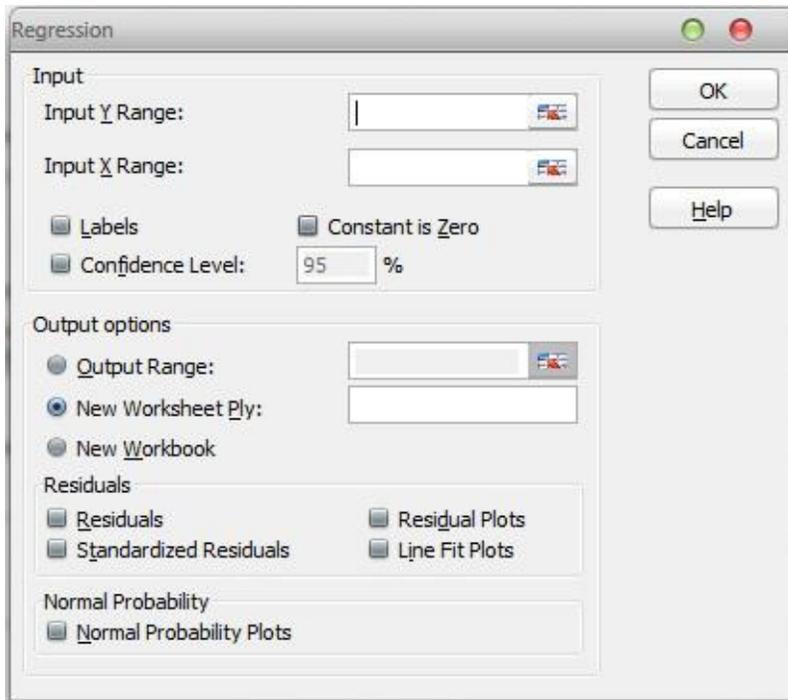
Once the Add-in is installed, create a table of data similar to the following:

x	y
7	216
12	302
14	431
6	151
22	677
38	958
14	431
9	227
10	308
11	277

To perform the Regression analysis, select a cell and then click the 'Data' ribbon tab. Then double-click the Data Analysis section of the ribbon. This will present a popup similar to the following:



Select the Regression option and click the 'Ok' button. This will prompt another popup to enter the cell ranges for the data which will be analyzed:



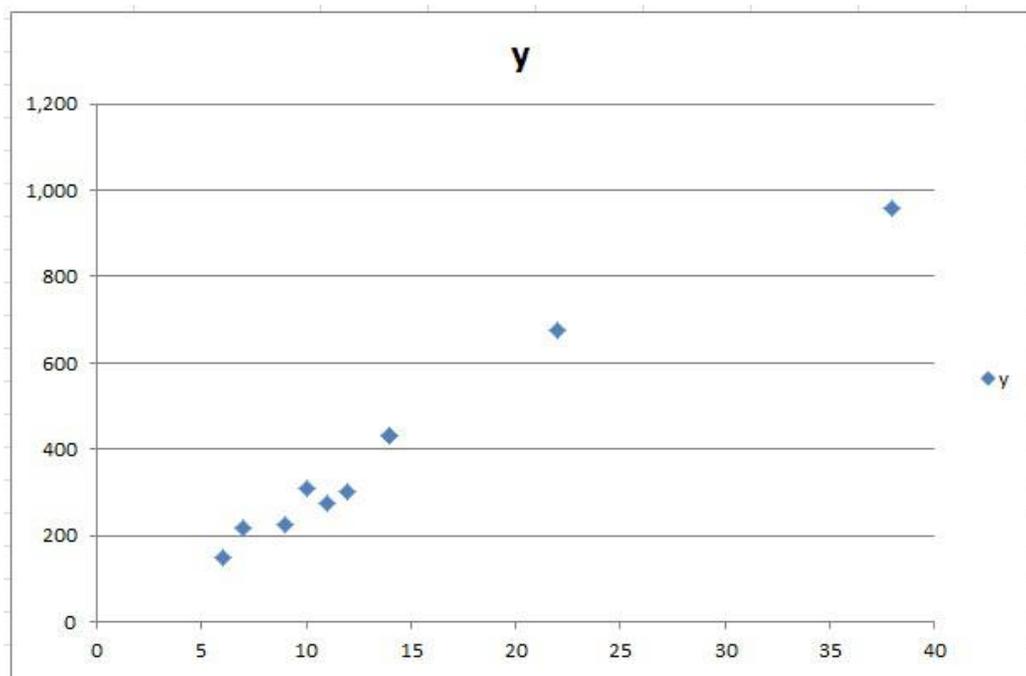
For the Input Y Range, select all cells in the Y column; including the header. Do the same for the Input X Range using the X column. Be sure to select the Labels checkbox and then click the 'Ok' button. Doing so will create the regression analysis within your spreadsheet:

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple F	0.983681							
R Square	0.967629							
Adjusted R Square	0.963582							
Standard Error	47.13044							
Observations	10							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	531179.4	531179.4	239.1323	3.04E-07			
Residual	8	17770.22	2221.278					
Total	9	548949.6						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	30.71877	28.02887	1.095969	0.304985	-33.9159	95.35346	-33.9159	95.353462
x	25.67002	1.659996	15.4639	3.04E-07	21.84206	29.49797	21.84206	29.497973

We focus on the R Square value, and the Coefficients of the rows labeled 'Intercept' and 'x'. The best way to demonstrate the significance of these values is to present them in a scatter chart with the linear trend line being rendered. This can be done in the following steps:

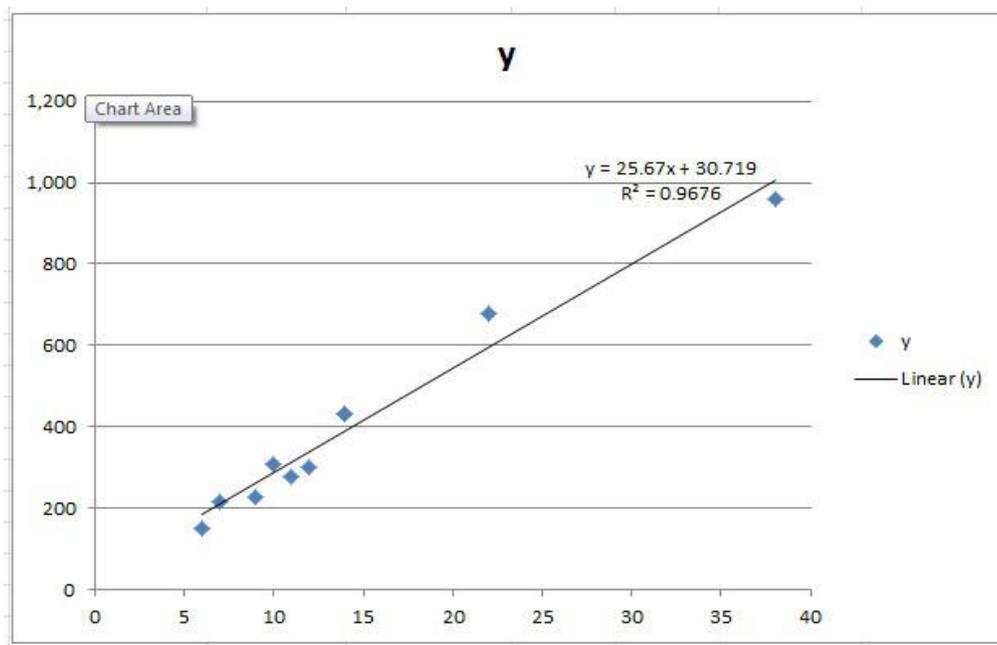
- 1) Click an empty cell
- 2) Select the 'Insert' ribbon tab
- 3) Select the Scatter Plot and choose 'Scatter with only Markers'
- 4) Right Click in the center of the Chart and choose 'Select Data'
- 5) Highlight your data table including the column header lines
- 6) Click the 'Ok' button

Your chart should be displayed in a way similar to the following:



The following steps can be used to display the linear trend line in the chart:

- 1) Left Click any of the plot markers
- 2) Right Click and select 'Add Trendline'
- 3) Select Linear in the Trendline Options pop up
- 4) Enable the checkbox for 'Display Equation on chart'
- 5) Enable the checkbox for 'Display R-Squared value on chart'
- 6) Click the close button



### Determining uncertainties from a linear regression

The uncertainty at the 95% confidence level can also be obtained from a linear regression. Two uncertainties can be obtained: (1) the uncertainty in the intercept (for a line  $y = mx + b$ , this would be the uncertainty in  $b$ ), (2) the uncertainty in the slope (i.e. in  $m$ ). Both will be described for the example in the “SUMMARY OUTPUT” block of data given earlier.

The uncertainties are obtained from the bottom section of the “SUMMARY OUTPUT” where lines beginning “Intercept” and “x” are found. Taking “Intercept” as the first example, the following two entries are labeled “Coefficient” and “Standard Error,” respectively. The “Coefficient” is the intercept itself ( $b$ ). The “Standard Error” is the equivalent of  $S$  discussed earlier. The uncertainty can be determined from the “Standard Error” if you know  $\nu$  (number of degrees of freedom) and the critical  $t$  value. However, it is easier to use the columns labeled “Lower 95%” and “Upper 95%.” There are two pairs of such columns; use the first pair. The uncertainty is just half the difference between “Upper 95%” and “Lower 95%.” In the example, then, the uncertainty in  $b$  is given by:

$$\Delta = \frac{1}{2}(95.35 - (-33.92)) = 64.64$$

In most cases, of course, this value would be rounded to  $\Delta = 60$ .

The uncertainty in the slope is determined from the line beginning “x”. The column labeled “Coefficient” is just the value of the slope from the fit (note that the slope,  $m$ , is by definition the *coefficient* of  $x$  in the equation of a line). Once again, use the values under “Upper 95%” and “Lower 95%” to determine the uncertainty:

$$\Delta = \frac{1}{2}(29.50 - 21.84) = 3.83$$

In most cases, of course, this value would be rounded to  $\Delta = 4$ .

And there you have it; Linear Regression done simply in Excel.

### APPENDIX - Fictitious data set

You may use the data set given below to test if you are using the regression software add-on for EXCEL correctly. If you are, you should obtain the same results for the slope, intercept, and 95 % confidence limits for the slope and intercept for the data presented here. Proceed as follows:

1) Enter the following data exactly as given below.

x	y
2.0	5.31
3.6	7.18
4.9	10.33
5.3	11.86
6.1	13.05

2) Use the linear regression software package you have added to EXCEL, and find the values for the slope and intercept, and the 95 % confidence limits on those values. If you have run the software correctly, you should get the following results:

intercept =  $a_0 = b = 0.873 \pm 2.9$  at 95% confidence limits

slope =  $a_1 = m = 1.980 \pm 0.63$  at 95% confidence limits

3) Confirm that you have found the correct slope, intercept, and 95 % confidence limits by comparing your results to those given above. Once you have done this, you can use the same software and method to analyze any experimental data you wish to fit to a straight line.