Curve Fitting (or how can we determine the equation when we have data points?)

CASE A: Let's say you have the following data points

X	y
0	0
6	26
7.8	34.1
8.4	37
9.2	39.7
12	49.9

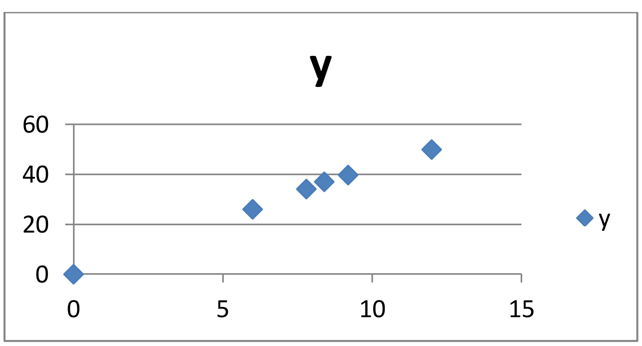
You want to find the relationship between y and x. That is you want to find the equation which, when you put in an X will give you the Y.

Steps:

1. Enter the data into an Excel spreadsheet

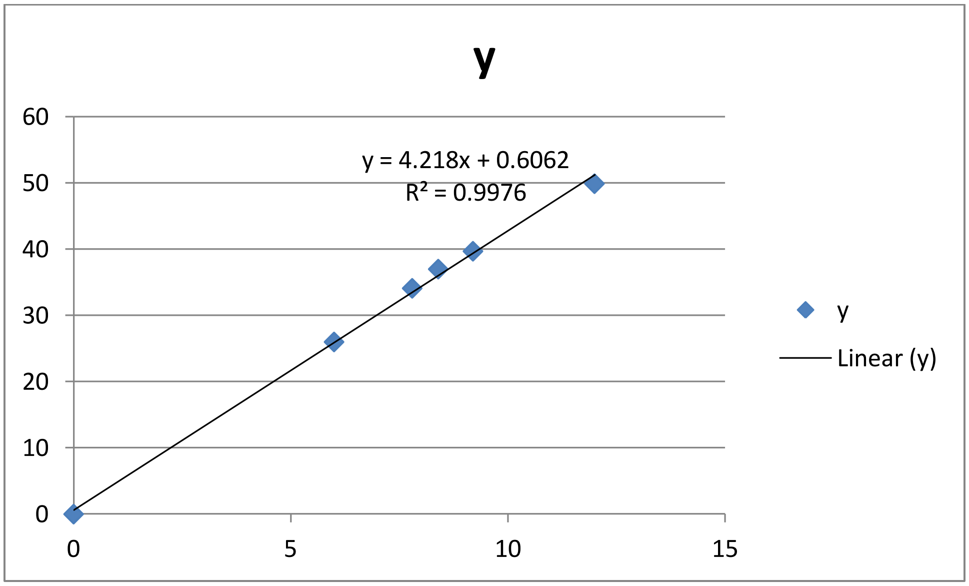
2. Highlight the data (including the X and Y at the top

3. Click insert>scatter chart (points, no line). That will give you a graph with the points, but no line connecting the points it will look like this:



4. Now we want to draw the line and get the equation. With the chart highlighted, click on layout> trendline> more trendline options.

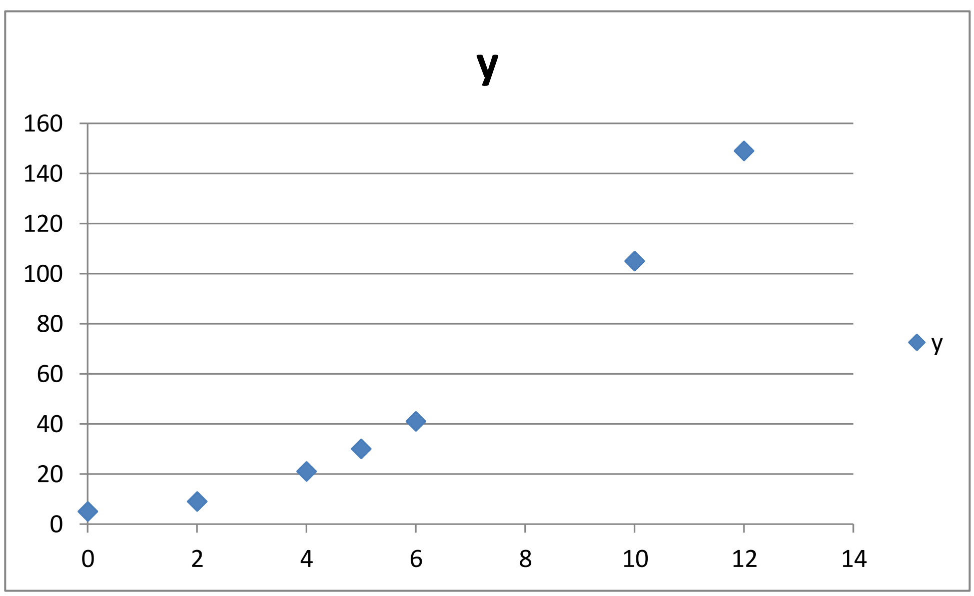
5. This data set looks like a straight line so we will use the linear fit. Also click on "display equation" and "show R squared value". Now the result will look like this:



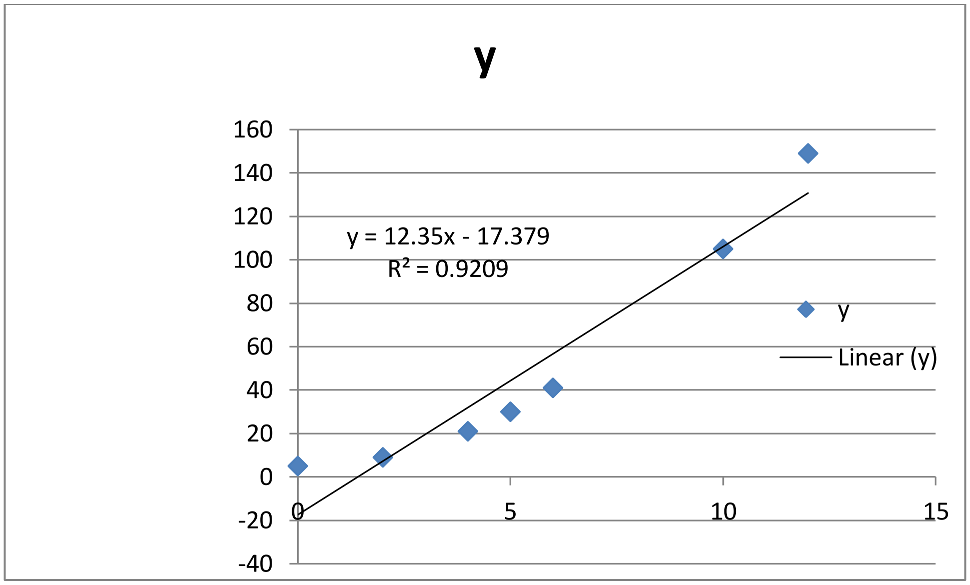
6. The line fits the data very nicely. The equation is y=4.218x + .6062 Clearly, the slope is 4.218 and the intercept is .6062 This is ignoring any consideration of significant figures.

CASE B: Of course, not every data set will be linear. Here is another data set and the scatter chart:

x	y
0	5
2	9
4	21
5	30
6	41
10	105
12	149

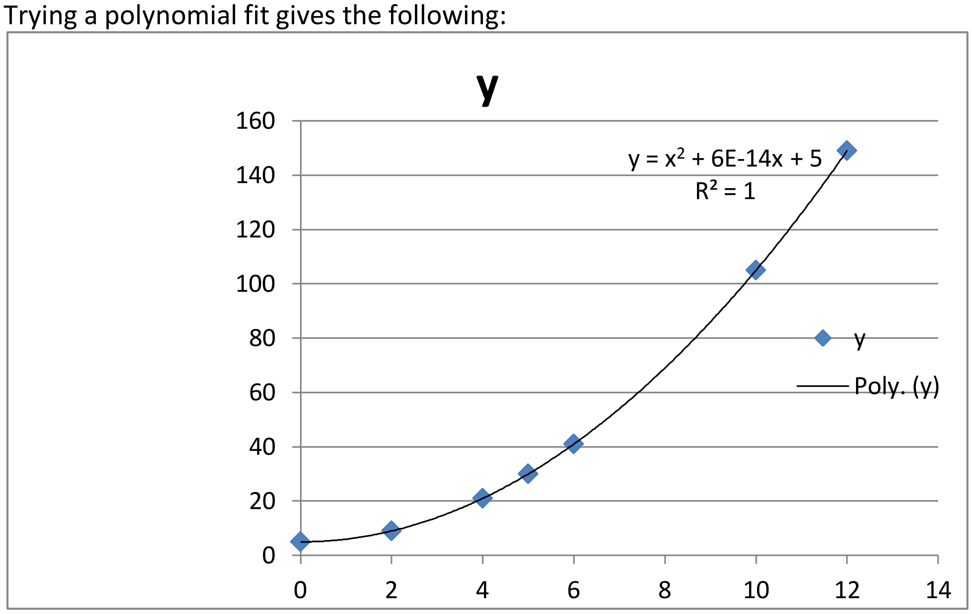


If I try a linear fit it looks like this:



Clearly that is not good.

Trying a polynomial fit gives the following:



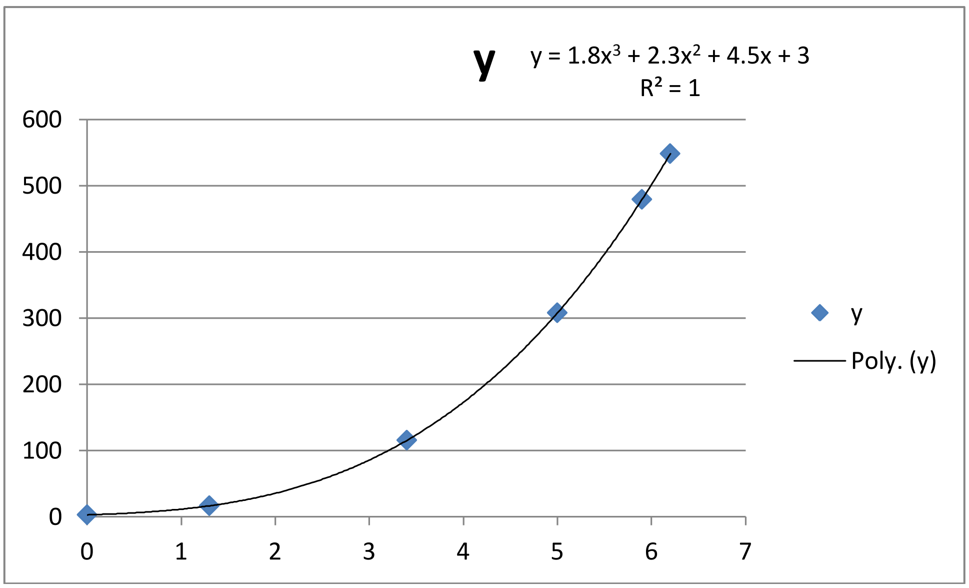
Realizing that 6E-14 is ridiculously small, we can drop it. The equation would be y = x^2 + 5

Note: R squared is a measure of the 'goodness' of the fit. The closer R squared is to 1, the better the fit. Look back at the previous graphs and see how R squared reflects the goodness of the fit.

Case C: Sometimes the equation is more complicated. Here is a completed graph with an equation of the form y = ax^3 + bx^2 +cx + d

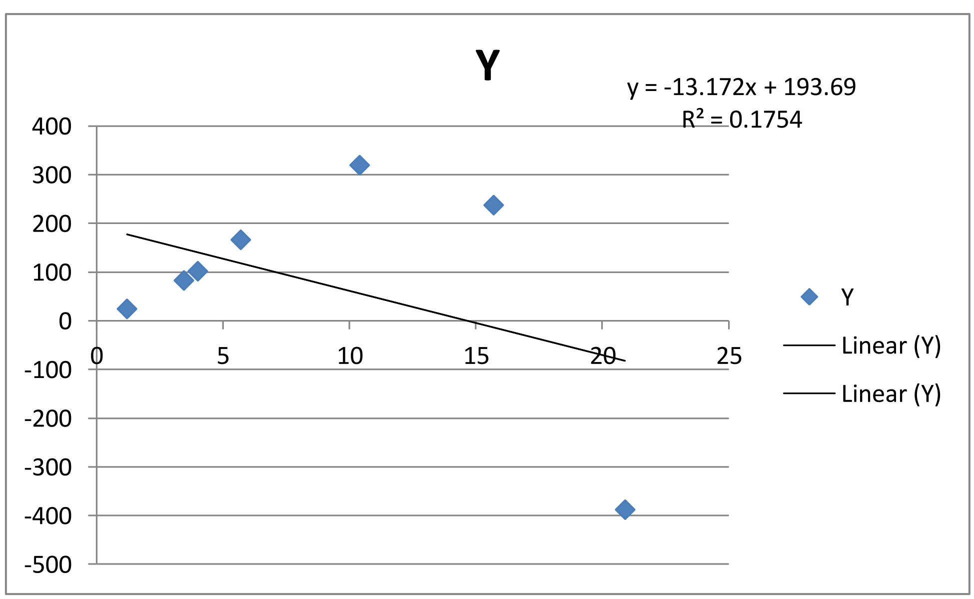
x	y
0	3
1.3	16.6916
3.4	115.6352
5	308
5.9	479.2952
6.2	548.3024

In this case I used the polynomial fit of order 3.



Case D: Sometimes a graph will even change directions. In this case it is obvious that a linear fit will be lousy. Here it is:

x	Y
1.2	24.85248
3.45	83.25337
4	102.24
5.7	166.7744
10.4	320.1062
15.7	237.9764
20.9	-387.512



Now that same data set with a polynomial fit order 4. Since the 2E-14 coefficient is so small it can be ignored yielding the equation $y = -0.34x^3 + 6x^2 + 4x + 12$

x	Y
1.2	24.85248
3.45	83.25337
4	102.24
5.7	166.7744
10.4	320.1062
15.7	237.9764
20.9	-387.512