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

<table>
<thead>
<tr>
<th>X</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>7.8</td>
<td>34.1</td>
</tr>
<tr>
<td>8.4</td>
<td>37</td>
</tr>
<tr>
<td>9.2</td>
<td>39.7</td>
</tr>
<tr>
<td>12</td>
<td>49.9</td>
</tr>
</tbody>
</table>

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:

![Graph showing linear fit with equation y = 4.218x + 0.6062 and R² = 0.9976]

6. The line fits the data very nicely. The equation is \( y = 4.218x + 0.6062 \) Clearly, the slope is 4.218 and the intercept is 0.6062. 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:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td>10</td>
<td>105</td>
</tr>
<tr>
<td>12</td>
<td>149</td>
</tr>
</tbody>
</table>

![Scatter chart showing data points]

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: $R^2$ squared is a measure of the ‘goodness’ of the fit. The closer $R^2$ squared is to 1, the better the fit. Look back at the previous graphs and see how $R^2$ 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 $y = ax^3 + bx^2 + cx + d$

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>16.6916</td>
</tr>
<tr>
<td>3.4</td>
<td>115.6352</td>
</tr>
<tr>
<td>5</td>
<td>308</td>
</tr>
<tr>
<td>5.9</td>
<td>479.2952</td>
</tr>
<tr>
<td>6.2</td>
<td>548.3024</td>
</tr>
</tbody>
</table>

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

![Graph](image)

**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:

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>24.85248</td>
</tr>
<tr>
<td>3.45</td>
<td>83.25337</td>
</tr>
<tr>
<td>4</td>
<td>102.24</td>
</tr>
<tr>
<td>5.7</td>
<td>166.7744</td>
</tr>
<tr>
<td>10.4</td>
<td>320.1062</td>
</tr>
<tr>
<td>15.7</td>
<td>237.9764</td>
</tr>
<tr>
<td>20.9</td>
<td>-387.512</td>
</tr>
</tbody>
</table>

![Graph](image)
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 \( y = -0.34x^3 + 6x^2 + 4x + 12 \)

<table>
<thead>
<tr>
<th>x</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>24.85248</td>
</tr>
<tr>
<td>3.45</td>
<td>83.25337</td>
</tr>
<tr>
<td>4</td>
<td>102.24</td>
</tr>
<tr>
<td>5.7</td>
<td>166.7744</td>
</tr>
<tr>
<td>10.4</td>
<td>320.1062</td>
</tr>
<tr>
<td>15.7</td>
<td>237.9764</td>
</tr>
<tr>
<td>20.9</td>
<td>-387.512</td>
</tr>
</tbody>
</table>