

HOW TO OBTAIN THE BEST-FIT EQUATION USING LINEAR REGRESSION (LINEAR LEAST SQUARES)

February 12, 2019

WHAT PROGRAM TO USE

Typically, there are three options that students use:

- A graphing calculator. If you have one and know how to use it, you're all set.
- Excel.
- Google Sheets.

The choice of between the latter two is mostly a matter of convenience. Each have their own advantages, but both can perform all of the tasks that you need for General Chemistry Lab.

BASIC DATA ENTRY

- 1) It is simplest if the x-values and y-values are in adjacent columns. The x-values should be in the first column and the y-values in the second column.

By this time, you should know that when we say graph Alpha vs. Beta that Alpha goes on the y-axis and Beta goes on the x-axis. We always plot y vs. x. Reversing your axes is a particularly good way to get incorrect values for your slope and y-intercept.

- 2) To enter data in a cell, move the cursor to that cell.
 - a. If the data you enter only contains numerals and no letters, then the spreadsheet treats it as a number. Exception: E (or e) can be used to indicate exponential notation. Do not use x or *. For example, if you want 1.5×10^{-3} , enter either $1.5E-3$, $1.5e-3$, or 0.0015 .
 - b. If the data you enter contains any letter (other than e), the spreadsheet interprets it as a label (string) and it will not be used for calculations.
 - i. When constructing a graph, if you enter labels in the top row of each column, the program will use the labels to identify the data (and Google Sheets will use it to label the axes and title the graph).
 - ii. If you enter a label in any row besides the first, at best the program will ignore that row. Even worse, it will usually treat the entire column of numbers as labels. This can cause the plot to have two sets of points and the linear regression values you obtain will be nonsensical. In Excel, labels are left-justified and numbers are right-justified in each cell, so it helps you see if you accidentally entered a label where you meant to enter a number (most frequently, this happens when a capital O is typed instead of a zero, 0).

Note: It is not necessary to add the column labels. All of the directions work the same way whether or not the column labels are included. However, we encourage the use of labels. It helps avoid accidentally switching your axes (which leads to wrong answers, frustration, and wasted time).

GRAPH CONSTRUCTION

Begin by entering your data. As an example:

Enter the following x-value data into cells A1 to A6: length, 0.8, 2.1, 3.4, 4.7, 6.0

Enter the following y-value data into cells B1 to B6: mass, 1.5, 4.2, 6.9, 9.3, 11.9

As noted above, adding a row containing the labels (length and mass) is optional.

It should look like this:

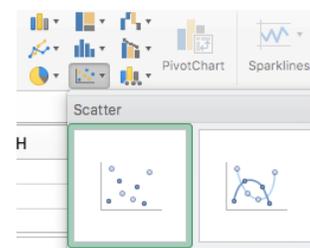
	A	B
1	length	mass
2	0.8	1.5
3	2.1	4.2
4	3.4	6.9
5	4.7	9.3
6	6	11.9

To construct a graph and perform linear regression on the data using **Google Sheets**:

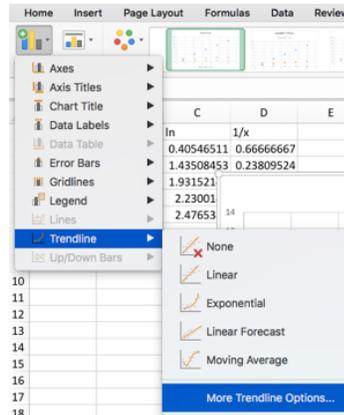
- 1) Select the entire data range (but NOT the labels): A2 to B6, in our example.
- 2) Select the **INSERT** menu found in the upper left and then choose **Chart**. A bar graph will now appear, with labeled axis and an appropriate title (“mass vs. length”) and the Chart editor will open.
- 3) Within the Chart editor, select the **Chart type** and then scroll down and select **Scatter chart**. (The icon looks like the one shown at the bottom of this page.)
- 4) The Chart Editor should open on the right side of the screen. Make sure that “**Use Row 1 as Headers**” and “**Use Column A as labels**” are both checked.
- 5) Select the **CUSTOMIZE** menu at the top of the Chart editor.
- 6) Select **Horizontal axis**. Make sure that “**Treat labels as text**” is NOT checked.
- 7) Select **Series**. Then select **Trendline**.
- 8) More options will appear. Move down to the **Label** scroll box and change it from **None** to **Use Equation**. The best fit line should appear on the graph and the equation will appear in the graph’s legend. You should see $y = 1.99x + -0.0138$. So, the slope is 1.99 and the y-intercept is -0.0138.

To construct a graph and perform linear regression on the data using **Excel**:

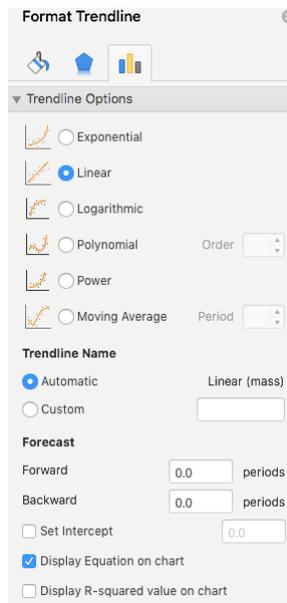
- 1) Select the entire data range (including the labels): A1 to B6, in our example.
- 2) Select the **INSERT** tab found in the upper left and then look for a set of icons that look like graphs near the middle of the ribbon. Choose X Y (Scatter) (this is the icon which just shows individual points. It will give you options for how to connect the points, but just choose the simplest option where they are not connected. The graph will now appear (with unlabeled axis and the title will be “mass”).



- 3) You should see that two new tabs have appeared at the top of the window: **CHART DESIGN** and **FORMAT**.
- 4) On the **CHART DESIGN** tab, select **Add Chart Element** on the left-hand side of the ribbon, and hover the mouse pointer over **Trendline**.
- 5) A new box will appear. Select **More Trendline Options...** found at the bottom of the menu.



- 6) The **Format Trendline** box will appear on the right. Make sure the **Trendline Options** tab is selected.



- 7) It should have **Linear** selected (by default). Select the box in front of **Display Equation on Chart**. You should see $y = 1.9923x - 0.0138$ appear for our example. So, the slope is 1.9923 and the y-intercept is -0.0138.

In either program, there are alternate keystrokes that can be used to achieve the same result (and there are additional steps that can be taken to beautify the graph, if so desired).