1-2023

Introduction to 2-d Plots with Excel

Maria Ballester
Nova Southeastern University, mballest@nova.edu

Victor Castro Ph.D.
Nova Southeastern University, castvict@nova.edu

Recommended Citation

Ballester, Maria and Castro, Victor Ph.D., "Introduction to 2-d Plots with Excel" (2023). *Biophysical Chemistry Laboratory Manual*. 1.
https://nsuworks.nova.edu/biophysical_chemistry_lab/1

Biophysical Chemistry Laboratory

Introduction to 2-d plots with Excel

Dr. Maria Ballester and Dr. Victor Castro

1 Basics

Scientists rely on programing tools to visualize and analyze numerical data. Microsoft Excel offers many features that makes it very useful for undergraduate data analysis. When you open Excel, you will be greeted by an interface depicted in Fig. 1. The main features are:

- The Quick Access toolbar: offers buttons for the most commonly used commands; you can customize it with whatever buttons you like.
- The Ribbon: it groups small icons for common tasks together in tabs, such as charts, tables, clip art or a hyperlink. By default, it is divided into eight tabs, with an optional ninth one (Developer).

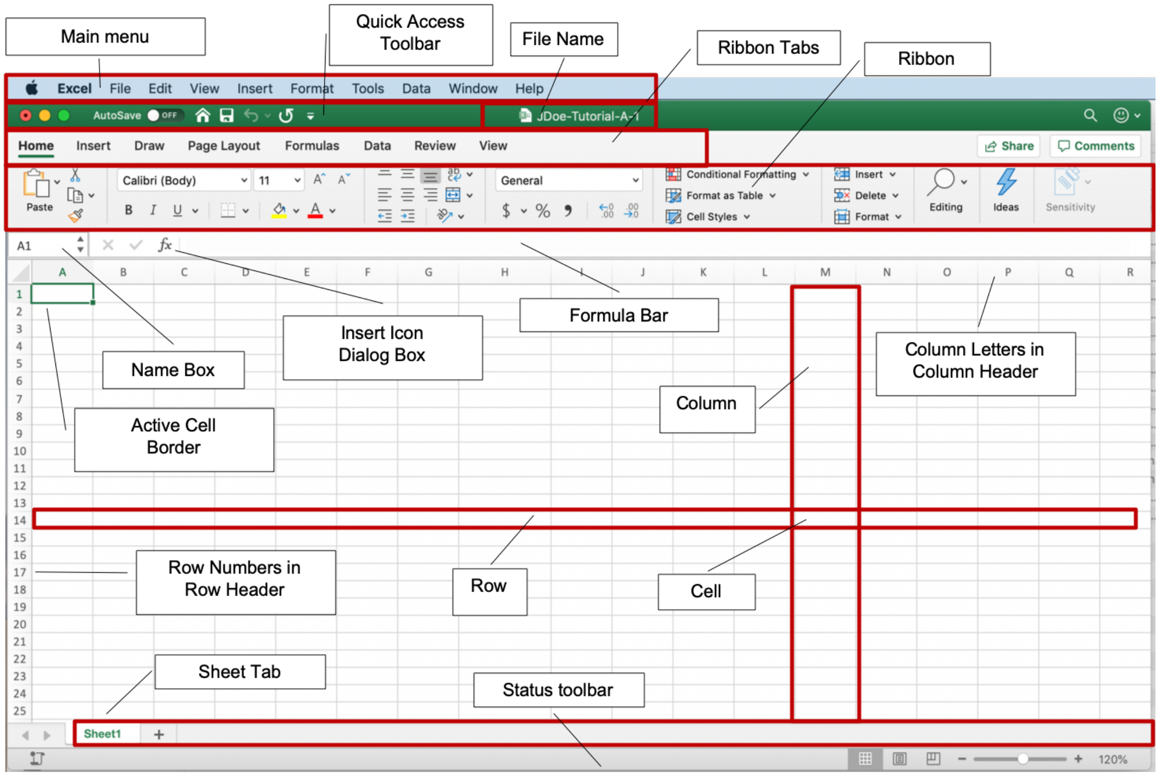


Figure 1. Excel 365's interface.

Excel creates a file called a *workbook* (or *book*) made of a number of *sheets*. Data goes into *cells*, that are defined by three parameters: (1) the sheet, (2) the column (labeled with letters) and (3) the row (labeled with numbers). When you open a new file, it will default to `Sheet1`, cell column `A`, cell row `1`. Figure 1 shows cell **`A1`** boxed in green, which you can see identified by its address just above the spreadsheet on the left; `Sheet1` is also highlighted on the bottom of the screen.

You can highlight an entire row or column by clicking on the letter or number, at the start of it. You can highlight specific number of cells by clicking *inside* a cell and dragging the mouse (or trackpad). When you highlight a cell (or groups of cells) you can also format its content, for example you can left justify or boldface it, or change the type of data it holds. Pressing `ENTER` moves you down a column. Pressing `TAB` moves you across a row. You can also move through cells by using the arrow keys.

You can use the tabs and menu options, or learn a few useful key combination (Table 1)

Shortcut	Description
`CMD+b`	Bold the selection
`CTRL+space`	Select entire column
`Shift+space`	Select entire row
`CMD+c`	Copy
`CMD+x`	Cut
`CMD+v`	Paste
`CMD+z`	Undo

Table 1. Excel shortcut. Both keys have to be pressed simultaneously. `CMD` and `CTRL` are the `command` and `control` keys.

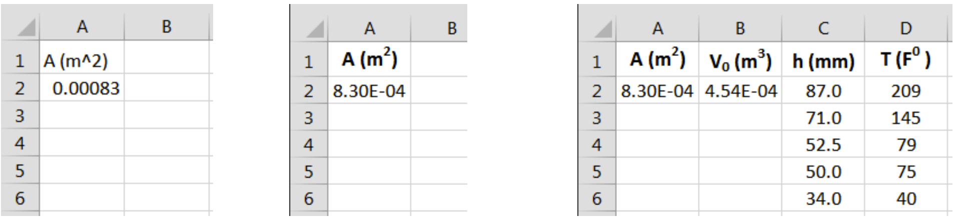
2 Data Manipulation

We are now going to practice how to enter data and manipulated it though formulas and/or functions.

Exercise 1

In this exercise you are going to enter the data obtained from an experiment that measured an area, an initial volume, and various heights and temperatures. You will use columns `A`, `B`, `C` and `D` for the data.

1) Create a new Workbook and use `Sheet1`.
2) Click on cell `A1` and type a label describing the experimental data; in this case "`A`" for *Area*. In science, table headings (and graph axes) should always be presented as "*Name of parameter* (*Units*)", hence you can use "`A (m^2)`", meaning that you will input data of an *area* in *meters squared*. Hit `ENTER` when you are done. The cell box will automatically move to the next box down the `A` column.
3) In cell `A2` type your first data point as shown in Fig. 2a, and then hit `ENTER`.
4) Now you can format their contents. Click on `A1` to select it and drag the mouse until `A2` is also selected. Choose *center justification*, on the *Home* tab, *Alignment* menu, to center the contents of both cells. From now on <*Tab Name*>→<*Menu Name*> will be used to help you find the right menu, e.g., *Home*→*Alignment*.
5) Select `A2` and choose the *Scientific* number option (*Home*→*Number*) menu. The arrows in this menu allows you to change the number of significant figures. Three is ok for this exercise.
6) Select `A1` and click on *Boldface* (*Home*→*Font* or `CMD+b`). **TIP**: You can change the label to "`A (m`^{2}`)`" by using *superscripts*. First click on the `A1` cell; in the *Formula bar* select and erase the "`^`" symbol. Now select the number `2`, right click on it, and from the list of options that appear, select *Format*. Under the *Effects* options of the *Format Cells Font* tab, tick *Superscript* and then click OK. Column `A` should now look like that shown in Fig. 2b. This technique can also be used for *subscripts*.
7) Repeat these procedures to enter the next data columns shown in Fig. 2c. Use columns `B` for the initial volume (V_0), `C` for height (h), and `D` for temperature (T). Your spreadsheet should now look like that shown in Fig. 2c.



	A	B
1	A (m^2)	
2	0.00083	
3		
4		
5		
6		

	A	B
1	**A (m^2)**	
2	8.30E-04	
3		
4		
5		
6		

	A	B	C	D
1	**A (m^2)**	**V_0 (m^3)**	**h (mm)**	**T (F^0)**
2	8.30E-04	4.54E-04	87.0	209
3			71.0	145
4			52.5	79
5			50.0	75
6			34.0	40

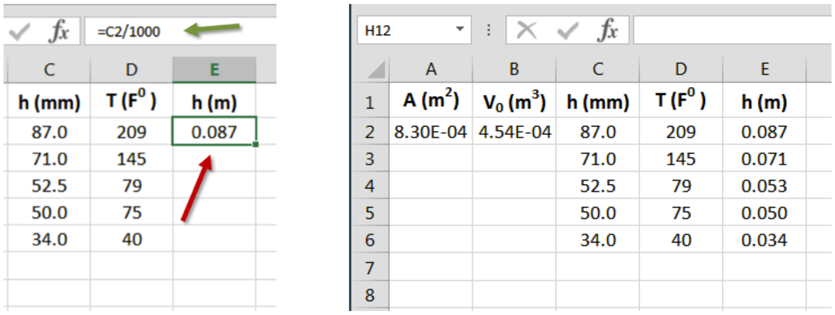
(a) (b) (c)

Figure 2

Exercise 2

You will now learn how to manipulate data in Excel.

1) Using the data from the previous exercise you will now change the units of all values in column `C`, from millimeters (mm) to meters (m), and place them in column `E`. The conversion factor you will use is $1\,\mathrm{m} = 1000\,\mathrm{mm}$, hence you will need to divide the values in `C` by `1000`. Label cell `E1` as "`h (m)`".



=C2/1000

C	D	E
h (mm)	T (F^0)	h (m)
87.0	209	0.087
71.0	145	
52.5	79	
50.0	75	
34.0	40	

(a)

H12

	A	B	C	D	E
1	A (m^2)	V$_0$ (m^3)	h (mm)	T (F^0)	h (m)
2	8.30E-04	4.54E-04	87.0	209	0.087
3			71.0	145	0.071
4			52.5	79	0.053
5			50.0	75	0.050
6			34.0	40	0.034
7					
8					

(b)

Figure 3

2) In cell `E2` type "`=`" then click on cell `C2` (do not press `ENTER` yet). Cell `E2` should now contain the phrase "`=C2`". The equal sign tells Excel that the text following is part of a *formula*, a *function*, or a combination of both. Complete the formula by typing "`/1000`" so that cell `E2` reads

```
=C2/1000
```

In this case we are telling Excel to take the contents of cell `C2` and divide it by `1000`. Now press `ENTER` and cell `E2` should display `0.087`.

Note: clicking on the cell displays the actual formula in the text box located in the toolbar above the spreadsheet (Fig. 3a green arrow), while the result of the calculation remains visible in the cell itself.

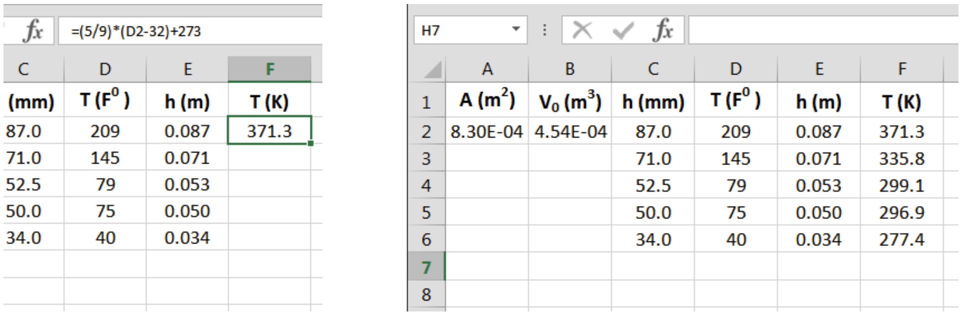
3) You can fill the remaining cells with the same formula by either (a) click on `E2` and copy (*Home→Copy* or `CMD+c`), select rows `E3` to `E6`, and click paste (*Home→Paste* or `CMD+v`); (b) select cell `E2`, position the cursor over the little square box at the lower-right corner of the cell (Fig. 3a, red arrow) and drag it down to cell `E6`; or (c) select cell `E2`, position the cursor over the little square box at the lower-right corner of the cell and double-click it. Column `E` should now look like that shown in Fig. 3b.

4) In the same way, use column `F` to change the units of all values in column `D` from Fahrenheit (F) to kelvin (K); you will need to use the formula $T_K = \frac{5}{9}(T_F - 32) + 273$. Label cell `F1` as "`T (K)`" and in cell `F2` (Fig. 4a), type in the formula

```
=(5/9)*(D2-32)+273
```

Now press `ENTER`. You can now fill the remaining cells with the same formula, as you did in step (3).

5) Your spreadsheet should now look like that shown in Fig. 4b.



=(5/9)*(D2-32)+273

C	D	E	F
(mm)	T (F^0)	h (m)	T (K)
87.0	209	0.087	371.3
71.0	145	0.071	
52.5	79	0.053	
50.0	75	0.050	
34.0	40	0.034	

(a)

H7

	A	B	C	D	E	F
1	A (m^2)	V$_0$ (m^3)	h (mm)	T (F^0)	h (m)	T (K)
2	8.30E-04	4.54E-04	87.0	209	0.087	371.3
3			71.0	145	0.071	335.8
4			52.5	79	0.053	299.1
5			50.0	75	0.050	296.9
6			34.0	40	0.034	277.4
7						
8						

(b)

Figure 4

In your previous exercises, the cell references in the formulas were **relative references**. This is why copying the formula on column `E` automatically resulted in successive values on column `C` being used to calculate the new values. On the other hand, if you have a formula that uses a **constant** you need to use an **absolute reference**, i.e., a reference that always points to the same cell. Excel uses a `$` prefix to denote absolute references. You can specify that both the row and column positions are absolute, or that only the row or the column are (Table 1).

Cell reference	Meaning
`$E$10`	always refers to column E row 10
`C$6`	always refers to row 6
`$M7`	always refers to column M

Table 1

Exercise 3

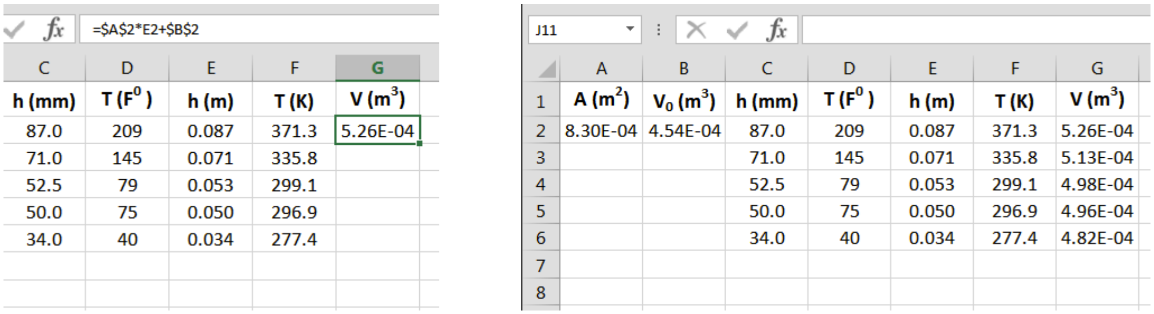
Excel handles complicated mathematical operations with built-in **functions**. In this exercise, you will find the average of the volume-to-temperature ratio in the data above.

1) In column G we will write a formula for find the volume of a circular cylinder with an initial value, given by: $V = A \cdot h + V_0$. Select cell `G1` an insert the label “$\mathtt{V\ (m^3)}$”. In cell `G2` (Fig. 5a) type:

```
=$A$2*E2+$B$2
```

In this formula, `A2` (area) and `B2` (initial volume) are **absolute references**, while `E2` is a **relative reference**.

2) Fill the series down in column `G` to match the series of *h* values in column `E`. Your spreadsheet should now look like that shown in Fig. 5b.



=$A$2*E2+$B$2

C	D	E	F	G
h (mm)	T (F^0)	h (m)	T (K)	V (m^3)
87.0	209	0.087	371.3	5.26E-04
71.0	145	0.071	335.8	
52.5	79	0.053	299.1	
50.0	75	0.050	296.9	
34.0	40	0.034	277.4	

(a)

J11

	A	B	C	D	E	F	G
1	A (m^2)	V$_0$ (m^3)	h (mm)	T (F^0)	h (m)	T (K)	V (m^3)
2	8.30E-04	4.54E-04	87.0	209	0.087	371.3	5.26E-04
3			71.0	145	0.071	335.8	5.13E-04
4			52.5	79	0.053	299.1	4.98E-04
5			50.0	75	0.050	296.9	4.96E-04
6			34.0	40	0.034	277.4	4.82E-04
7							
8							

(b)

Figure 5

3) Select cell `H1` (Fig. 6a). Insert the label “$\mathtt{V\ (m^3)/T\ (K)}$”, and in cell `H2`, type the formula:

```
=G2/F2
```

4) Fill in for the other rows in column **`H`** with the formula.

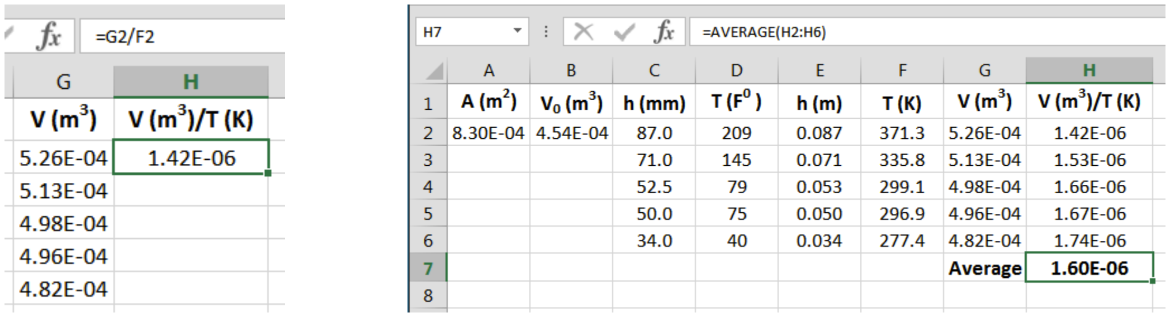
5) In cell `G7` (Fig 6b) type in the label “`Average`”, and in cell `H7`, type the built-in function

```
=AVERAGE(H2:H6)
```

This means that Excel will find the average of the values between `H2` and `H6` (the range is referred as `H2:H6`).

6) To account for significant figures, change column `H` so its values have 3 significant figures.

7) Your spreadsheet should now look like that shown in Fig. 6b. Print out the results of exercise 3.



`=G2/F2`

G	H
V (m^3)	**V (m^3)/T (K)**
5.26E-04	1.42E-06
5.13E-04	
4.98E-04	
4.96E-04	
4.82E-04	

(a)

H7 `=AVERAGE(H2:H6)`

	A	B	C	D	E	F	G	H
1	**A (m^2)**	**V_0 (m^3)**	**h (mm)**	**T (F^0)**	**h (m)**	**T (K)**	**V (m^3)**	**V (m^3)/T (K)**
2	8.30E-04	4.54E-04	87.0	209	0.087	371.3	5.26E-04	1.42E-06
3			71.0	145	0.071	335.8	5.13E-04	1.53E-06
4			52.5	79	0.053	299.1	4.98E-04	1.66E-06
5			50.0	75	0.050	296.9	4.96E-04	1.67E-06
6			34.0	40	0.034	277.4	4.82E-04	1.74E-06
7							**Average**	**1.60E-06**
8								

(b)

Figure 6

3 Making 2-D graphs

3.1 Single data set

Graphs form an integral part of science. In Excel, graphs are made of at least two blocks: the *chart area*, and the *plot area*. You can add other blocks, such as legends, labels, and textboxes, which can be placed anywhere within the chart area. The data are shown in the plot area. The chart area takes up the total space of the graph, and it contains the plot area plus the regions surrounding it. Axis labels and legends can be either inside or outside the plot area but must be within the chart area. We are going to illustrate in the following example.

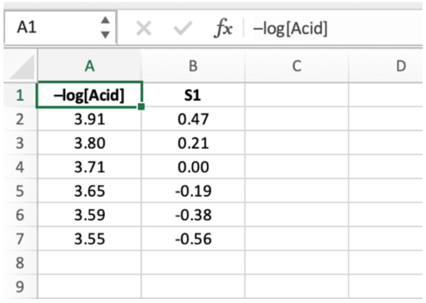
Consider the case of a freebase porphyrin of concentration [X] that has been protonated by the process of adding an acid gradually. The relation between the concentration of the system components can been described in terms of acid–base equilibrium: $\mathrm{P}^{2+} \rightleftarrows 2\mathrm{H}^{+} + \mathrm{P}$. A spectral analysis of the process will give us information about the absorbance of each and the maxima peaks. Then, the dissociation constant can be evaluated graphically by a logarithmic form similar to the Henderson–Hasselbalch equation, modified to include the absorbance of the species using

$$\log_{10}\left(\frac{A - A_{\mathrm{p}}}{A_{\mathrm{fb}} - A}\right) = -\mathrm{p}K_{a,\mathrm{obs}} - \log_{10}[\mathrm{Acid}]$$

were $[\mathrm{Acid}]$ is the concentration of the acid added to the solution. We are going to plot $\log_{10}(A - A_{\mathrm{p}})/(A_{\mathrm{fb}} - A)$ vs. $-\log_{10}[\mathrm{Acid}]$ and later use it to find the apparent value of $\mathrm{p}K_a$.

Exercise 4

1) Start a new spreadsheet. Copy the data from Fig. 7 in the range `A1:B1` to `A7:B7`. Notice that in cell `A1` we wrote “`–log[Acid]`”, where “–“ is an “en dash” and not the negative sign (-). Make the column’s label boldface.



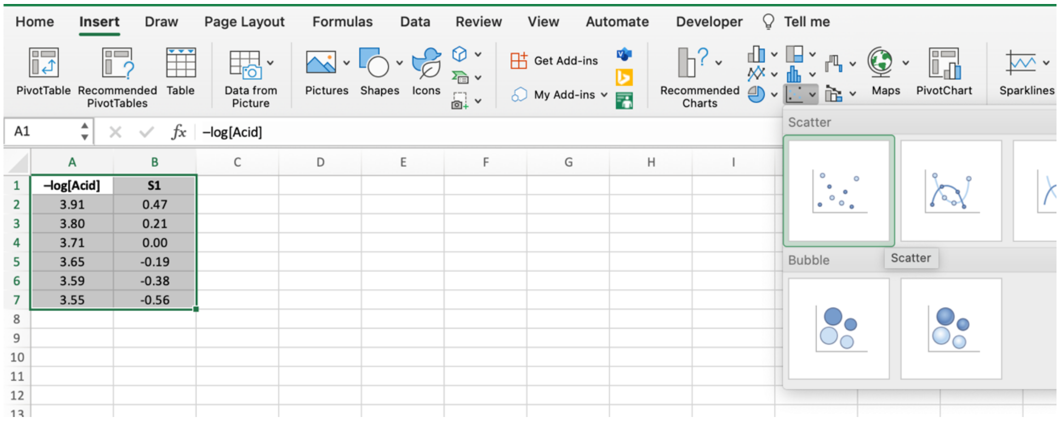
A1 `–log[Acid]`

	A	B	C	D
1	**–log[Acid]**	**S1**		
2	3.91	0.47		
3	3.80	0.21		
4	3.71	0.00		
5	3.65	-0.19		
6	3.59	-0.38		
7	3.55	-0.56		
8				
9				

Figure 7

The first column represents the logarithm of the concentration of the acid during the process, and the second one is the $\boldsymbol{\log_{10}(A - A_{\mathrm{p}})/(A_{\mathrm{fb}} - A)}$ of the solution at the wavelength where the freebase porphyrin has its maximum peak (soret or S1).

2) Select the two-column block `A1:B27`. With your data highlighted, select the *Insert* tab. Under the *Charts* there are several options for different types of graphs. You are going to create a **scatterplot**, which plots the x and y −coordinates of each data point with additional colored lines, bars, etc. Click on the *Scatter* option (Fig. 8).



A1: −log[Acid]

	A	B
1	−log[Acid]	S1
2	3.91	0.47
3	3.80	0.21
4	3.71	0.00
5	3.65	-0.19
6	3.59	-0.38
7	3.55	-0.56

Figure 8

3) At this point you have a rough draft version of your graph (Fig. 9). This default graph has many missing features that you need to fix (e.g., unlabeled axes, gridlines, legend, etc.). You can now move the graph anywhere on the Excel window by clicking around the edge and dragging it.

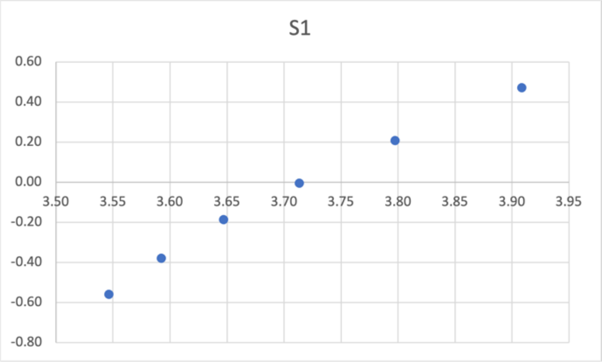
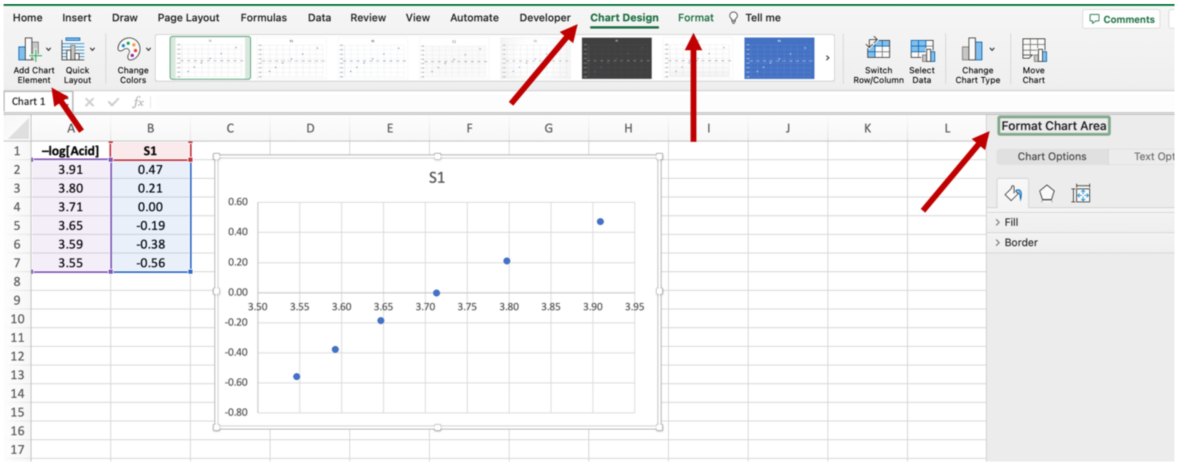


Figure 9

4) You need to edit the graph further by **double-clicking o**n it. Notice the two new Ribbon-tabs: *Chart Design*, and *Format*, referred to as *Chart Tools*. You can also see the *Formatting Task Pane* that just opened on the right side of the Excel window.



	A	B
1	−log[Acid]	S1
2	3.91	0.47
3	3.80	0.21
4	3.71	0.00
5	3.65	-0.19
6	3.59	-0.38
7	3.55	-0.56

Figure 10

5) Editing the graph requires that you use the formatting task pane, and the *Chart Design→Add Chart Element* menu (check red arrows in Fig. 10).

6) The *Formatting Task Pane* stays visible until you close it by clicking the × button. The fastest way to open it is by double-clicking a chart element. You can also use: (a) the keyboard shortcut `CTRL+1` while a chart element is selected, (b) select an element, and then on then select on the ribbon *Format*, or (c) on a chart, right-click on an element, and then select *Format* <element> where <element> is the axis, series, legend, title, or area that was selected.

7) Let's modify the plot. First, click on the "Chart Title" and delete it.

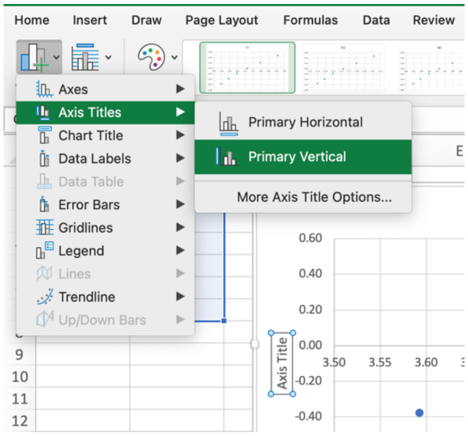
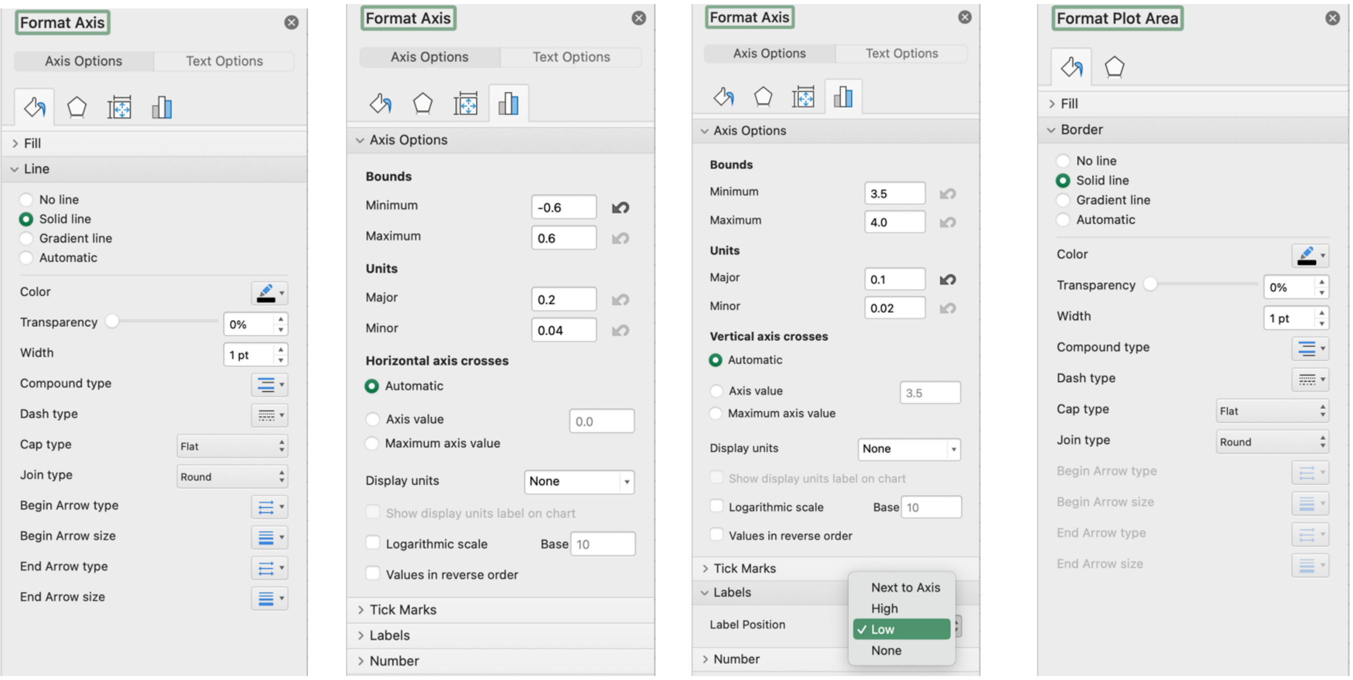


Figure 11

8) You need to add labels to the two axes. Select *Chart Design→Add Chart Element* menu (Fig. 10) and click on *Axis Titles* and set both axes. Now click inside each axis title: write "`log(A-Ap)/(Afb-A)`" for the y −axis, and "`–log[Acid]`" for the x −axis.



(a) Format Axis — Line: ○ No line ● Solid line ○ Gradient line ○ Automatic; Color; Transparency 0%; Width 1 pt; Compound type; Dash type; Cap type Flat; Join type Round; Begin Arrow type; Begin Arrow size; End Arrow type; End Arrow size

(b) Format Axis — Axis Options: Bounds: Minimum -0.6, Maximum 0.6; Units: Major 0.2, Minor 0.04; Horizontal axis crosses: ● Automatic ○ Axis value 0.0 ○ Maximum axis value; Display units None; ☐ Show display units label on chart; ☐ Logarithmic scale Base 10; ☐ Values in reverse order

(c) Format Axis — Axis Options: Bounds: Minimum 3.5, Maximum 4.0; Units: Major 0.1, Minor 0.02; Vertical axis crosses: ● Automatic ○ Axis value 3.5 ○ Maximum axis value; Display units None; ☐ Show display units label on chart; ☐ Logarithmic scale Base 10; ☐ Values in reverse order; Labels: Label Position: Next to Axis, High, ✓ Low, None

(d) Format Plot Area — Border: ○ No line ● Solid line ○ Gradient line ○ Automatic; Color; Transparency 0%; Width 1 pt; Compound type; Dash type; Cap type Flat; Join type Round; Begin Arrow type; Begin Arrow size; End Arrow type; End Arrow size

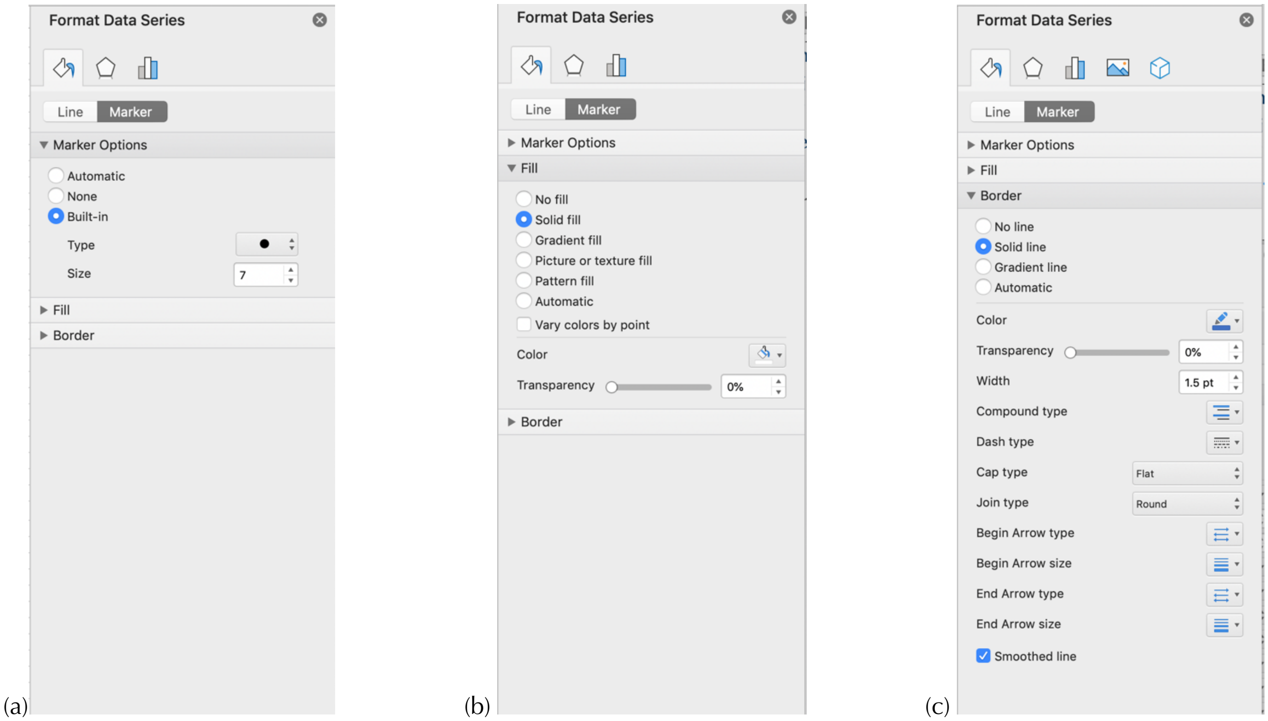
Figure 12

9) Since Excel automatically determines the size, color, numbering, etc. of each axis, you now need to edit these settings. You will begin by double-clicking on the x −axis to pull up the *Format Axis* menu. First set the line color of the x −axis to black, with a width of `1 pt` (Fig. 12a).

10) Next, change the *Minimum* and *Maximum* values on the x −axis; and the *Major Unit* to the values in Fig. 12b.

11) Repeat for the *y*-axis (Fig. 12c). Make sure to set the *Label Position* to *Low*. That way the numbers are out of the way.

12) The next step is to Put a border around your data by double-clicking on the *Plot Area*. Chose *Border* and set the *Color* to black, and the *Width* to `1 pt`. (Fig. 12d)



(a) Format Data Series — Line | Marker — Marker Options: ○ Automatic, ○ None, ● Built-in; Type ●; Size 7; Fill; Border

(b) Format Data Series — Line | Marker — Marker Options; Fill: ○ No fill, ● Solid fill, ○ Gradient fill, ○ Picture or texture fill, ○ Pattern fill, ○ Automatic, ☐ Vary colors by point; Color; Transparency 0%; Border

(c) Format Data Series — Line | Marker — Marker Options; Fill; Border: ○ No line, ● Solid line, ○ Gradient line, ○ Automatic; Color; Transparency 0%; Width 1.5 pt; Compound type; Dash type; Cap type Flat; Join type Round; Begin Arrow type; Begin Arrow size; End Arrow type; End Arrow size; ☑ Smoothed line

Figure 13

13) Finally, you will use a better marker. Double-click on any of the blue points in the plot area to bring the *Format Data Series* task pane. Click on *Marker*, Set the *Marker Options* to *Built-in* and set the *Size* to 7. This is the place where you can change the shape of your marker and its size (Fig. 13a). In *Fill* set to *Solid fill* and chose white color (Fig. 13b); in *Border* set to *Solid line*, and *Width* to `1.5 pt`. (Fig. 13c).

14) Figure 14 shows how the end result should look. You can further increase its clarity by changing the font, and/or its size on your axes (and axes labels).

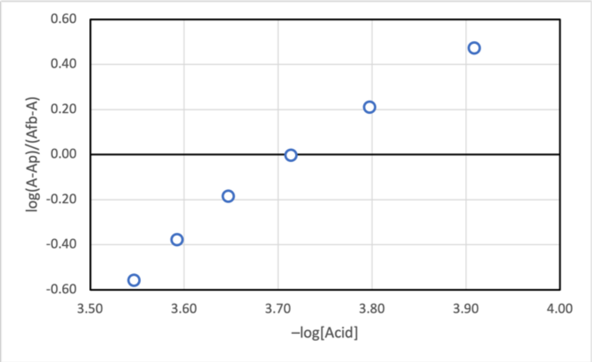


Figure 14

3.2 More than one data set on the same plot

Exercise 5

Now we are going to add and additional data series to our plot.

1) In cell `C1` type the label “`S2`”. From cell `C2:C7` copy: `0.33, 0.05, -0.19, -0.39, -0.64, -0.89.`
2) Highlight the data in column `C` and copy (**`cmd+c`**). Click on the plot border and paste the data into the graph.
3) As you did before, change the marker size and style, line thickness, and colors to distinguish the two curves. Also, change the *Minimum* and *Maximum* values on the y −axis.
4) As a final touch, Select *Chart Design→ Add Chart Element→ Legend→ Right.*
5) Your plot should now look similar to Fig. 15.

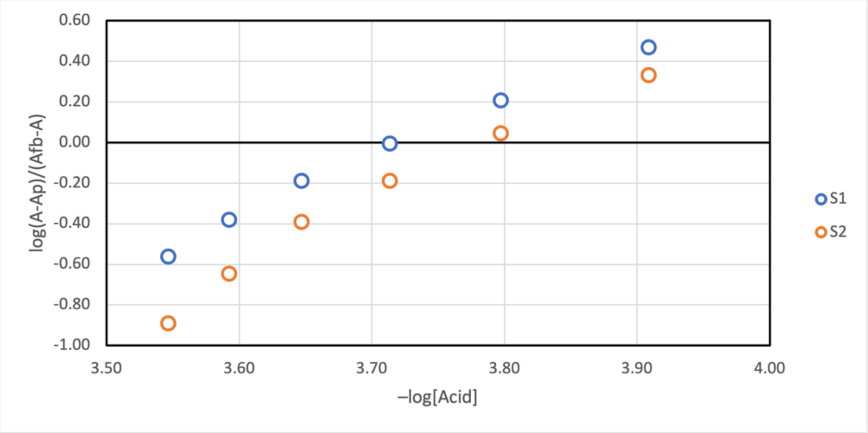


Figure 15. This plot shows the values of $\mathbf{log_{10}}(A - A_p)/(A_{fb} - A)$ as a function of $-\mathbf{log_{10}}[\mathbf{Acid}]$.

4 Plotting Scientific Data

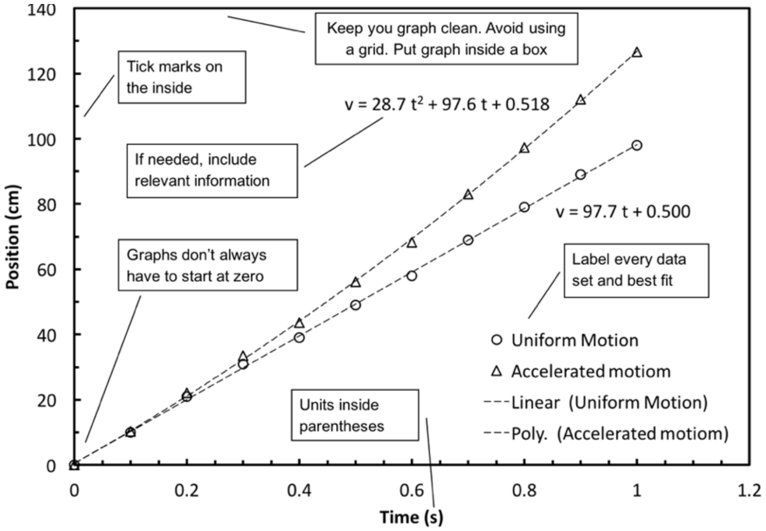


Figure 16. Requirements for a “good” data plot.

To make an Excel plot be an acceptable scientific plot here are a few simple guidelines:

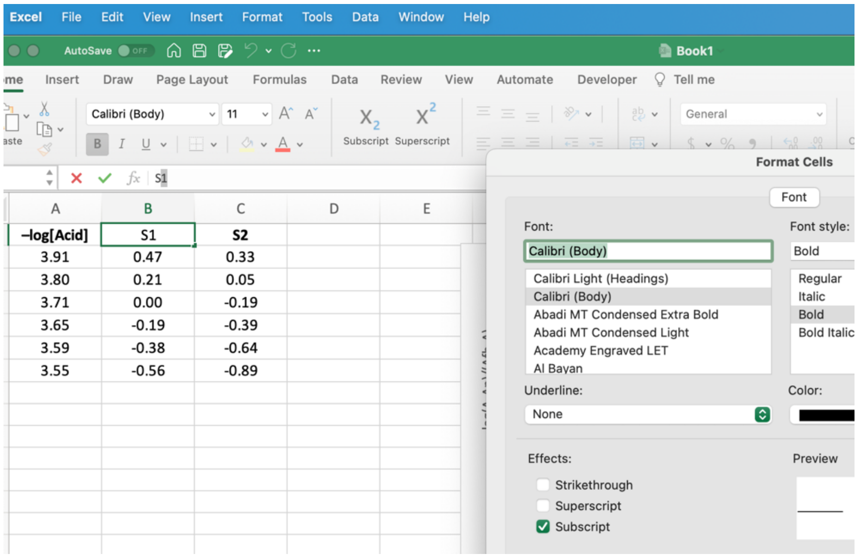
1. Graphs should **display data points, without connecting lines**. By connecting the points, you are assuming a relationship between the plotted data that could be false. An exception is when there are too many data points on the plot and individual data points cannot be distinguished. In this case no markers are used.

2. Caption every graph. This allows you to later recognize it, e.g., "Figure 3. Electric Field Inside a Parallel Capacitor." **Do not use** titles like "*y* vs. *t*", "Displacement versus time" or "Data from Table 1".
3. Make sure the axes are divided in multiples of 1, 2, or 5; this makes reading subdivisions easier.
4. The origin (zero) only needs to be included if it is important to the interpretation of the graph.
5. Label both axes. Set the proper quantity being measured and its units, e.g., "Absolute Pressure (kPa)".
6. If you have more than one set of data on a graph, be sure to label each set so you don't get confused later.
7. Use **markers** like ⊙ ⊡◇ or similar. In Fig. 16 circles were used to plot the data of "Uniform motion", and triangles for "Accelerated motion".

Exercise 6

With the plot you made in Exercise 5, follow the guidelines above to get a scientific plot of Fig 15.

1) As you did before, change the marker size and style, line thickness, and colors to distinguish the two curves. Also, change the *Minimum* and *Maximum* values on the y −axis.
2) As a final touch, you can use sub or superscripts where needed. Select the text to change. Go to *Format* → *Cells…*→ *Fonts*. Once the new window is open, you can select how to modify your text (Fig. 17).



A	B	C
−log[Acid]	S1	S2
3.91	0.47	0.33
3.80	0.21	0.05
3.71	0.00	-0.19
3.65	-0.19	-0.39
3.59	-0.38	-0.64
3.55	-0.56	-0.89

Font: Calibri (Body)

Font style: Bold

Underline: None

Effects:

☐ Strikethrough

☐ Superscript

☑ Subscript

Figure 17.

3) Your plot should now look similar to Fig. 18.

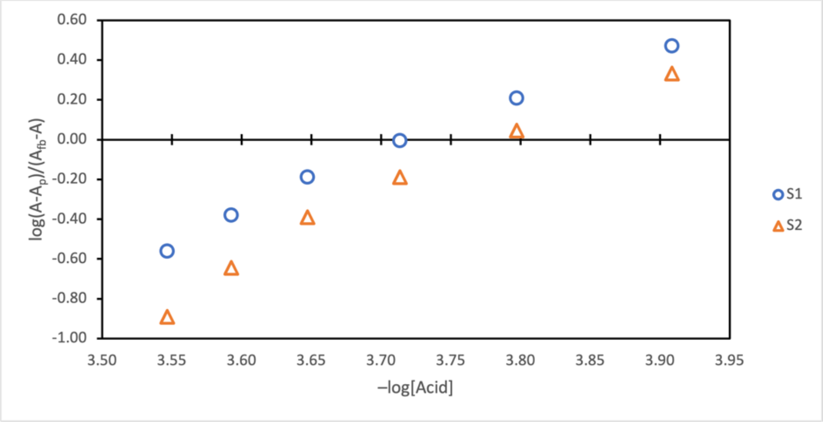


Figure 18.

5 Linear Regression

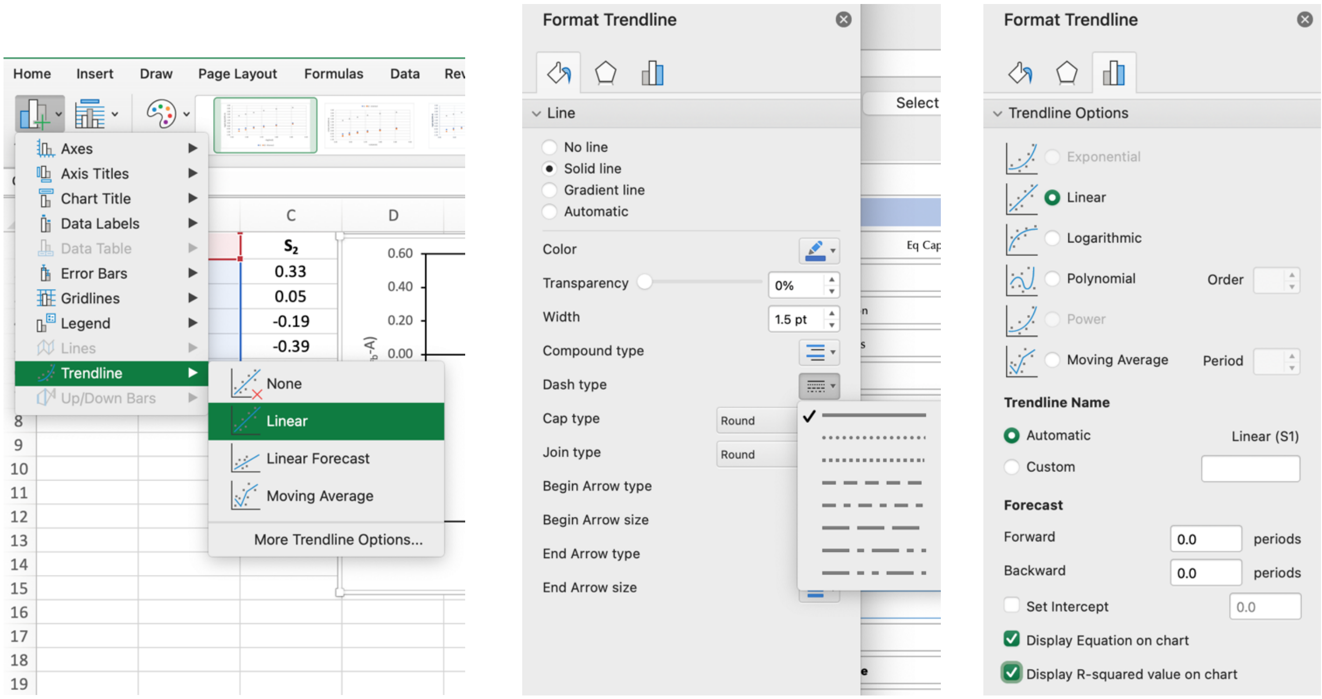
The goal of a linear regression analysis is to estimate the magnitude of the linear relationship between two variables and whether or not it is statistically significant. We are going to look at how to use the data we already plotted to perform a linear regression and how to interpret the results.

5.1 Plotting Trend lines

We now are going to use Excel's "trend-line" feature to illustrate the relationship (if it exists) between the variables. The plot of $\log_{10}(A - A_\mathrm{p})/(A_\mathrm{fb} - A)$ vs. $-\log_{10}[\mathrm{Acid}]$ is used to find the apparent value of $\mathrm{p}K_a$. First, we need to fit the experimental data to a straight line by using the least-squared linear method. Then, we use the fact that when the straight line intercepts the horizontal axis the value of $\log_{10}(A - A_\mathrm{p})/(A_\mathrm{fb} - A)$ is zero. Hence, we get $\mathrm{p}K_a = -\log_{10}[\mathrm{Acid}]$. This intercept point corresponds to $[\mathrm{P}]/[\mathrm{P}^{2+}] = 1$, indicating that the concentrations of the two species are equal. This is called the **isosbestic point**.

Exercise 7

1) Highlight the chart, click on the S2 data (triangles) and hit `Delete`. You should now have only one data set (S1).
2) Click to open the *Chart Tools→Design→ Add Chart Element* menu and select *Trendline* (Fig. 19a)
3) By default, Excel creates a linear best fit line in your plot. Double-click anywhere on the line to bring the *Format Trendline* task pane. Choose *Line* and change the *Color* to black and the *Dash type* to the one in Fig. 19b.
4) Now select *Trendline Options* and click on *Display Equation on chart*, and *Display R-squared value on chart*. Here is where you can set the type of best fit you want (Exponential, Linear, Logarithmic, etc.). (Fig. 19c).
5) You can move the equation text around and place it in a place that doesn't interfere with the plot. You can also edit the text to reflect the variables you are using, and not the generic x or y assumed by Excel.
6) Your plot should now look similar to Fig. 20.



Format Trendline — Line:
○ No line
● Solid line
○ Gradient line
○ Automatic
Color
Transparency 0%
Width 1.5 pt
Compound type
Dash type
Cap type Round
Join type Round
Begin Arrow type
Begin Arrow size
End Arrow type
End Arrow size

Format Trendline — Trendline Options:
○ Exponential
● Linear
○ Logarithmic
○ Polynomial
○ Power
○ Moving Average
Trendline Name
● Automatic Linear (S1)
○ Custom
Forecast
Forward 0.0 periods
Backward 0.0 periods
☐ Set Intercept 0.0
☑ Display Equation on chart
☑ Display R-squared value on chart

(a) (b) (c)

Figure 19

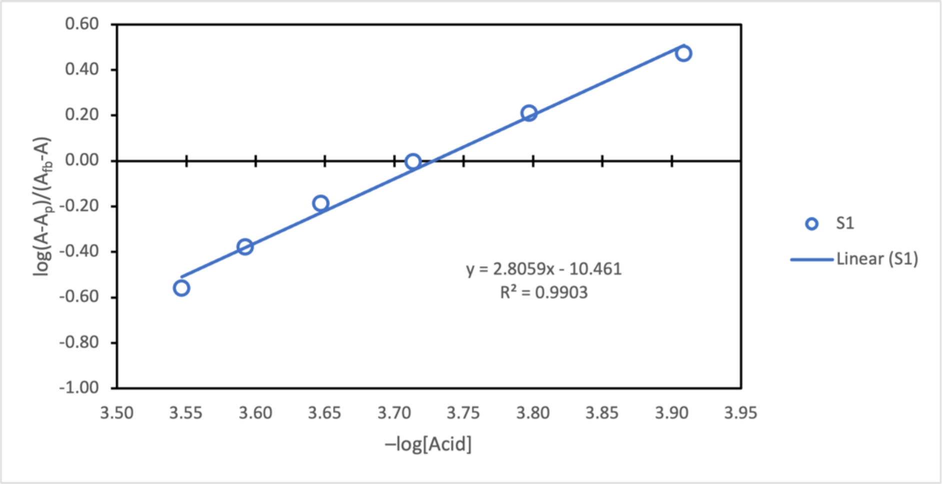


Figure 20. $\mathrm{p}K_a$ example.

7) We can now use the values in the plot to find the $\mathrm{p}K_a$ values. Copy the slope and the intercept values from the plot into cells `A9-A10` and `B9-B10`, as in Fig. 21. Write "`pKa`" in cell `C9`.

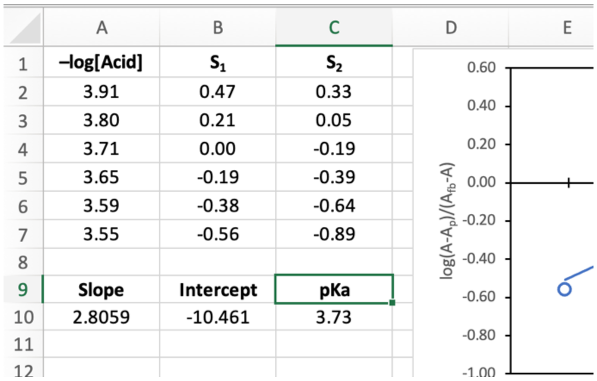
8) To find the place where the line intercepts the horizontal axis, we need to solve $y = m\,x + b$ for x when $y = 0$.

$$0 = m\,x + b \;\;\Rightarrow\;\; x = -b/m$$

Insert this formula into cell `C10`:

```
=-B10/A10
```

9) Your worksheet should now look similar to Fig. 21. Make sure you adjust the significant figures.



	A	B	C	D	E
1	**–log[Acid]**	**S_1**	**S_2**		
2	3.91	0.47	0.33		
3	3.80	0.21	0.05		
4	3.71	0.00	-0.19		
5	3.65	-0.19	-0.39		
6	3.59	-0.38	-0.64		
7	3.55	-0.56	-0.89		
8					
9	**Slope**	**Intercept**	**pKa**		
10	2.8059	-10.461	3.73		
11					
12					

Figure 21. $\mathrm{p}K_a$ value.

5.2 Excel's Built-In Functions

To avoid writing manually the slope and the y-intercept, we can use the following functions:

```
intercept(known_ys, known_xs)
slope(known_ys, known_xs)
```

where `known_ys` is the range of cells that contain the vertical values, and `known_xs` is the range of cells that contain the horizontal values.

Exercise 7

1) Click on an `A11` and enter

```
=slope(B2:B7,A2:A7)
```

2) Excel returns a more precise calculation for the slope (`2.80594706`).

3) Now, click on an `B11` and enter

```
=intercept(B2:B7,A2:A7)
```

4) Click on cell `C10` and copy. Paste it into cell `C11` to get a new value of pK_a. Adjust the significant figures and compare to our previous result (Table 2).

Slope	Intercept	pKa
2.8059	-10.461	3.7282
2.80594706	-10.460628	3.7280

Table 2.

If we need more statistical details, we can use the `LINEST` function.

```
=LINEST(known_ys,known_xs,const,stats)
```

where `known_ys` and `known_xs` is the data. **`const`** can have the values of `1` (true), where the intercept is treated normally or `0` (false), where the intercept is set to have the value of `0`. **`stats`** can have the values of `1` (true), where a set of regression statistics is returned with the slope and intercept or `0` (false), where the regression statistics are not returned.

Keep in mind that you will need enough room for this information (an area of 2 columns by 5 rows), as seen in Table 3.

Column 1	Column2
Slope	Intercept
Std. error for slope	Std. error for intercept
R2	Std. error for Y estimate
F statistics	Degrees of freedom
Regression sum of squares	Residual sum of squares

Table 3

Besides the slope and *y*-intercept, the other values are:

Coefficient of determination (R^2). it measures how well the regression model fits the data. Values range from 0 to 1, and a higher value indicates a better fit.

Standard errors. Generally, these values show the precision of the regression analysis. The smaller the numbers, the more certain you can be about your regression model.

F statistic. You use it to support or reject the null hypothesis. It is recommended to use the F statistic in combination with the P value when deciding if the overall results are significant.

Degrees of freedom. You can use the degrees of freedom to get F-critical values in a statistical table, and then compare the F-critical values to the F statistic to determine a confidence level for your model.

Regression sum of squares. It indicates how much of the variation in the dependent variable your regression model explains.

Residual sum of squares. It indicates how much of the variation in the dependent variable your model does not explain. The smaller the residual sum of squares compared with the total sum of squares, the better your regression model fits your data.

Exercise 7

1) Starting on row 16, use columns `D` and `G` to write the labels from Table 3 (Fig. 22)

2) Click on cell `E16` and type

```
=LINEST(B2:B7,A2:A7,1,1)
```

The output should appear in columns `E` and `F`, as shown in Fig. 22.

3) We can use this information not only to determine the $\mathrm{p}K_a$ value, but also its experimental error, by using

$$\delta(\mathrm{p}K_a) = \mathrm{p}K_a\sqrt{\left(\frac{\delta\ \text{slope}}{\text{slope}}\right)^2 + \left(\frac{\delta\ \text{intercept}}{\text{intercept}}\right)^2}$$

where δ slope is in cell `E17`, slope is in cell `E16`, δ intercept is in `F17`, and intercept is in `F16`.

4) Click on cell `D22` and write "`pKa value`". Also write "`pKa error`" on cell `D23`.

5) Click on cell `E22` and write the formula:

```
=-F16/E16
```

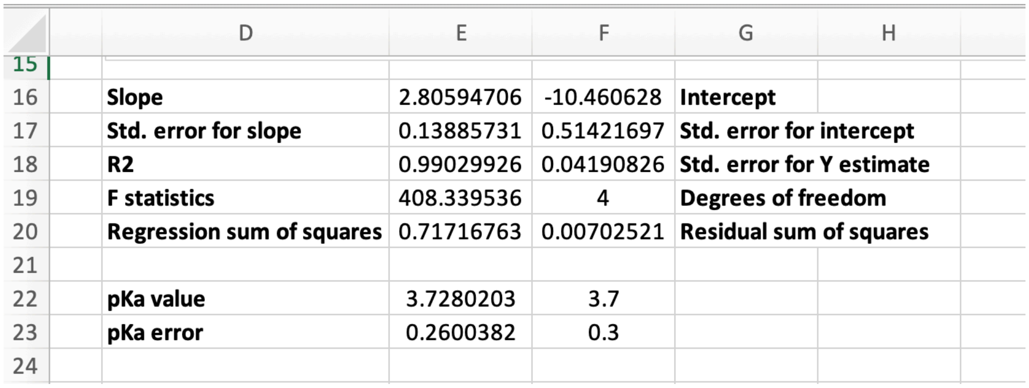
6) Write the error formula into `E23`

```
=E22*SQRT((E17/E16)^2+(F17/F16)^2)
```

7) One additional step is still needed when reporting an experimental value. The experimental error must have only one significant figure (two in very rare cases). Therefore, the results in E22 and E23 where copy to F22 and F23 and the significant figure adjusted. Then, the result for the $\mathrm{p}K_a$ value, according to the data is

$$\mathrm{p}K_a = 3.7 \pm 0.3$$

8) The final result is shown in Fig. 22.



	D	E	F	G	H
15					
16	**Slope**	2.80594706	-10.460628	**Intercept**	
17	**Std. error for slope**	0.13885731	0.51421697	**Std. error for intercept**	
18	**R2**	0.99029926	0.04190826	**Std. error for Y estimate**	
19	**F statistics**	408.339536	4	**Degrees of freedom**	
20	**Regression sum of squares**	0.71716763	0.00702521	**Residual sum of squares**	
21					
22	**pKa value**	3.7280203	3.7		
23	**pKa error**	0.2600382	0.3		
24					

Figure 22. LINEST output for the data of the example. The value and experimental error of $\mathrm{p}K_a$ is also shown.

6 References

1. E. Joseph Billo, *Excel for Chemists: A Comprehensive Guide*. John Wiley & Sons, Nov 1, 2011.

2. E. Joseph Billo, *Excel for Scientists and Engineers: Numerical Methods*. Wiley-Interscience; 1st Ed., April 6, 2007.

3. R. de Levie, *Advanced Excel for Scientific Data Analysis*, 2nd edition Oxford University Press; 2nd edition (August 14, 2008)

4. D. M. Bourg, *Excel Scientific and Engineering Cookbook: Adding Excel to Your Analysis Arsenal*. O'Reilly Media; 1st Ed., February 7, 2006.