

Introduction to Scientific Analysis of Data Using Spreadsheets.

Computer spreadsheets are very powerful tools that are widely used in Business, Science, and Engineering to perform calculations and record, present, and analyze data, as well as perform a variety of more sophisticated calculations that can use a number of special functions.

One of the virtues of spreadsheets is their relative ease of use compared to programs like Matlab and Mathematica or programming languages like C, or FORTRAN. In fact, spreadsheets have many of the functions of programming languages as built in functions and can do many of the same things. You can also make your own custom functions that can be called in spreadsheets if you need to, and Microsoft Visual Basic for Excel is a full- edged programming language (albeit with a bit of a learning curve).

We will be using Microsoft Excel for this lab since you'll be exploring some data analysis tools available. MS office suite is usually available completely free to all CSUF students, staff, and faculty, ask your instructor if you need help getting it. Note that while other spreadsheet software may have similar tools they have different names and methods and often require less secure extensions, so we'll focus on doing everything in excel. (Its good for your resume anyway).

Don't worry if you have never used Excel before, we will learn how to use the basic features of it in this lab then progress to more advanced features. You should become fairly proficient in using Excel by the end of the semester since we will use it to analyze and plot the data for almost all of the experiments.

Learn how to use Excel yourself. Don't let your partner be the one who always makes the spreadsheet, you should take turns working on it.

Note on Grading/Collection

Be sure your instructor tells you up front how they will be collecting/grading this lab.

This experiment is to show you the basics and prepare you for later experiments where we'll be using pre-created analysis sheets. Its up to your instructor if they want you to

- Assemble a report by copy pasting graphs and answering questions in a separate document to be submitted later.
- Inspect your final excel sheet at the end of lab to verify you have completed everything necessary.
- Hold you on the honor system until you use the sheets in a specific lab later and struggle with them since you didn't take the time now to figure out how to use excel.

Part I: Intro to Excel Data Analysis with Linear Graphs

Introduction

The basic linear equation

$$y = Ax + B$$

is the simplest and most widely used relation to data. Remember ...

- **A** is the slope
- **B** is the y intercept of the function.
- **x** is called the independent variable and
- **y** the dependent variable, since y depends on x.

The independent variable is the quantity that you directly vary at will in your experiment, like moving a ball from one height to another. The dependent variable is the quantity that is responding to your change, such as the potential energy of a ball that you are moving. When graphing the independent variable is usually put on the horizontal axis, and the dependent variable on the vertical axis.

Step 1: Graphing and Formatting

Purpose:

Often a visual representation of data can help us discover relationships between variables, or we may know what the relationship is and be attempting to test if our experiment verifies it. First though we need to review how to plot and show this data.

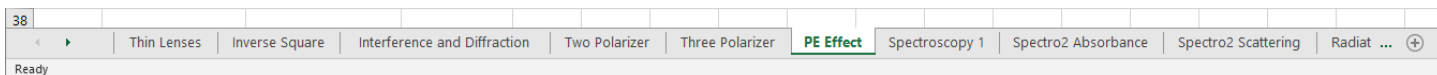
Instructions:

Let's take some data from a lab experiment on the photoelectric effect that we will perform later in the semester. This experiment involves shining a light source on a simple circuit which causes a current to flow in the circuit. The student adjusted the wavelength of the light source, and measured how much voltage it took applied against the current to stop the current from flowing (this is called the Stop voltage, or V_s)

Question 1:

In the experiment described above what is the independent variable and what is the dependent variable?

- A. First open the excel spreadsheet from the website, and go to the first tab on the bottom, "PE Effect". You may have to maximize your window to see the tabs.



You should see the data table shown in this image.

- B. Make a simple scatter plot of the data. Be sure the dependent variable is on the y axis, and independent on the x axis.

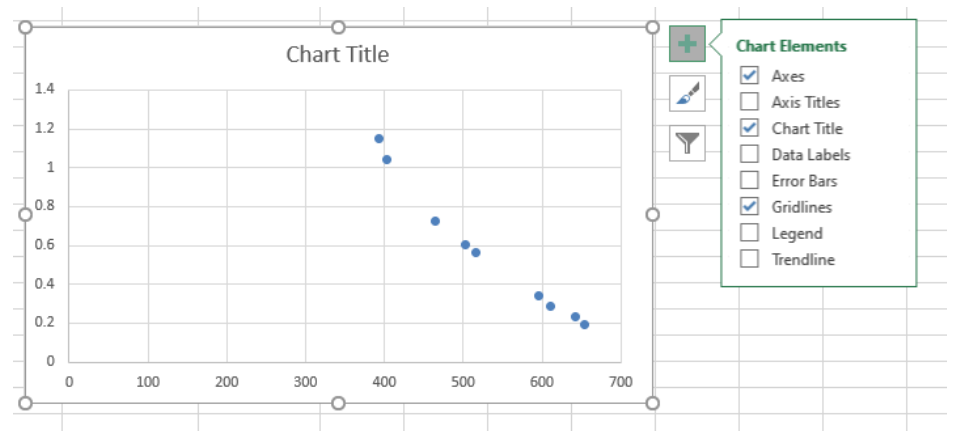
If they're reversed you can

- Click the POINTS on the graph then move the colored boxes
- Select the data in the opposite order by holding ctrl while you click and drack
- Right click the graph, go into select data, then manually select the x and y data sets.

Experimental Photoelectric Effect Data		Calcul
Wavelength λ (nm)	Stop Voltage (Volts)	f(Hz)
393.00	1.1500	
403.00	1.0440	
464.10	0.7270	
503.50	0.6070	
515.70	0.5630	
596.23	0.3394	
610.19	0.2894	
642.70	0.2307	
653.70	0.1924	

For all charts you make for the remainder of this class you should go through the following list. Make sure you know how to do every part of the list (and do them) before moving on to the next part.

Most can be adjusted through the green highlighted chart elements menu on the graph, or by right clicking whatever you want to change and opening its properties.



Excel's biggest user base is business, so default graph formats are mostly setup for that purpose. As a result you have to do a lot of formatting work to get a graph into a proper scientific or engineering format.

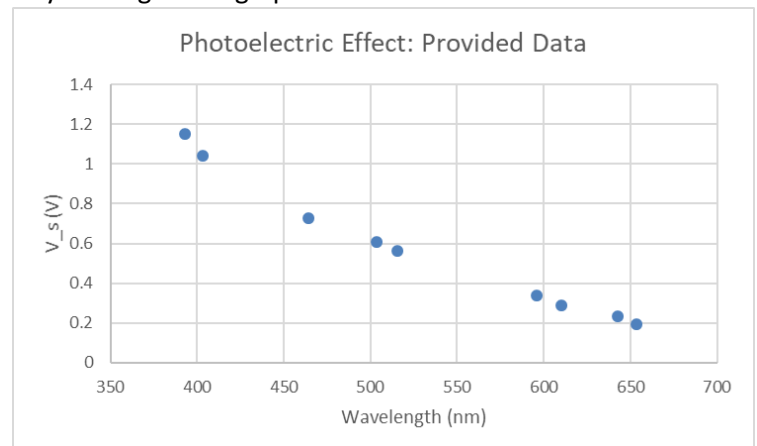
- C. Format the graph to match the standard below by changing:
 - a. Minimize white space
The data should occupy most of the graph window area, rescale your axes to fix this.
 - b. Title the chart
In scientific papers this is usually done with a figure caption, but for this lab each graph **MUST** have a chart title. The title should have information that gives the reader an idea of what is plotted.
 - c. Title the Axes
All Axes **MUST** have a title, and **MUST** include units if applicable. You need to know what data is being shown where.
 - d. Change legend titles
For this set there is only one series of data, so having the legend that says "series 1" in unnecessary. If there were multiple data sets to distinguish from this would be nice to leave on.
To edit these titles or right click your graph-> select data-> edit. The first box will be displayed in the legend.

In summary you are deleting unnecessary white space and information. Then you are making sure all necessary information is included so that you can tell what the data is by looking at the graph alone.

The graph should look similar to the one shown here. Though you have some freedom in exact titles.

Question 2:

Did you do all of the above? If yes then be sure to save your lab.



Step 2: Learning the Theory Behind the Collected and Graphed Data

Purpose:

The reason we're making this graph is to extract information from it through a theoretical "fit" of some kind.

By looking at the graph now you may see that its quite linear, but we need to go to the theory to find out whether or not the dependent variables response should be linear.

You'll see later in this lab that not all relationships are linear, and not all relationships can be easily fit by excel.

Theory:

Luckily in known cases the theory has often already been developed (the case everything for this class, but not necessarily once you get out to industry).

You'll eventually learn in your lecture that the photoelectric effect (what our above data set is for) is described by the following equation, (which earned Einstein the Nobel Prize in physics in 1921)

$$hf = e * V_s + \Phi, \quad (1)$$

Note that for our experiment our dependent variable was the stopping voltage, y , so we need to invert this equation so that the dependent variable is alone in the form...

$$y = Ax + B \quad (2)$$

dividing by e and solving for V_s in equation (1) gives,

$$V_s = \left(\frac{h}{e}\right) * f - \frac{\Phi}{e}; \quad (3)$$

Here h is Planck's constant, e is the electron charge, Φ is the work function of the surface, V_s the stopping voltage, and f is the frequency of the light. Don't worry too much about the variable or the theory here yet, you'll revisit this in a later lab.

Question 3:

Compare the equation of a line (2) to the stopping voltage equation for the PE Effect (3). What are the following equal to?

- $y =$
- $A =$
- $x =$
- $B =$

Equation (3) shows that if we plot the frequency of the light, not the wavelength as the independent variable and V_s as the dependent variable, we should get a straight line whose slope and intercept give us the information you found in **Q3**.

Since the data taken gives us V_s and λ we need some way to get frequency out of lambda. Luckily there's a relation between the wavelength of a light wave, and its frequency through the speed of light which is,

$$f = \frac{c}{\lambda} \quad (4)$$

where c is the speed of light given by $c = 2.9979E + 8$ m/s.

Step 3: Using Functions in Excel

Purpose:

Often when taking actual data in industry you may have hundreds to thousands of collected data points. Luckily the days of doing manual calculations on all these are over, we now have analysis tools to manipulate large data.

For this class we won't make you take any more than 30 data points, but even that many calculations is too many to efficiently do by hand. So here are the basics of using functions.

Note that functions in excel always start with an equals sign "=", this tells excel to evaluate what you type instead of just storing it as text.

Instructions:

Use a function in excel to add a frequency column to your data. This will allow you to easily convert an entire column of data, rather than calculate it one at a time.

- A.) The equation to convert to frequency is $f = c/\lambda$.

Enter this equation in cell D5 to calculate the frequency corresponding to the 393 nm wavelength. Check the equation bar in the picture here if you need help

- To tell excel you're entering an equation always start with an equals sign =, otherwise you're just entering text.
- To use a variable in a calculation just enter the location of the variable, in your sheet the first wavelength is in cell B5 and the speed of light is in cell E2.

Experimental Photoelectric Effect Data		Calculated Data
Wavelength λ (nm)	Stop Voltage (Volts)	f(Hz)
393.00	1.1500	=SE\$2/(B5/1E9)
403.00	1.0440	7.44E+14
464.10	0.7270	6.46E+14
503.50	0.6070	5.95E+14
515.70	0.5630	5.81E+14
596.23	0.3394	5.03E+14
610.19	0.2894	4.91E+14
642.70	0.2307	4.66E+14
653.70	0.1924	4.59E+14

- B.) Note: c is in m/s, but lambda is in nm (nanometers)... better fix that in your calculations. (notice the 1E9)
- C.) Now you don't want to rewrite this equation 10 times, click the bottom right corner of the cell (in the red circle) and drag it down or double click it.
- Notice that any normally entered variable will change in the direction you drag. Example dragging down the equation shown will change B5 to B6.
 - If you don't want it to change, such as when you have a reused value in each equation that may change you can add \$ before the column and row to fix the value. For example \$E\$2 does not change as you drag.
You can either manually type \$ in or use F4 on your keyboard after clicking into the variable.

Step 4: Extracting Data from a Graph “Fit”

- A.) Create a properly formatted graph of V vs Frequency
- B.) As in the picture above, add a linear trendline to your data and display the equation and R^2 values on the graph. There are many ways to add a trendline now...
 - a. Right click your data points then click add trendline, this should automatically open the Format Trendline option menu on the right.
 - b. In the “+” menu on the graph go to trendline, then click the arrow, then select more options to open the menu.
- C.) In the trendline options menu check the boxes for “Display Equation on chart” and “Display R-squared value on chart”
- D.) Right click the equation that has now appeared on your graph and select “format trendline label” and change the number format to “Scientific” with 4 decimal places.
 - a. You should now have a slope value of 3.0276E-15, or something very close to that.

Discussion

In experiments there are two types of error, systematic and random. This data is very precise, but is way off of the theoretical prediction in a predictable way.

Systematic Error

Systematic error is consistent difference between observed and true values of something.

The evidence for this error here is that our slope should reproduce the accepted value of h/e of 4.14E-15 Vs, but if you look back at that graph its about 30% off, which is consistently reproduceable in this experiment. When you perform the experiment, you may see the reason for this, having to do with reverse leakage currents.

Random Error

Random error is a chance difference between observed and true values, like a random scale misreading .

In the date you’re given there is very little random error which is what the good value for R^2 (> 0.99) indicates. Since this value is very close to 1, the data is of high quality and linear to high degree. It indicates that the experimenter likely took good data.

Coefficient of Determination:

In statistics, the coefficient of determination, denoted R^2 and pronounced R squared, indicates how well data points fit a line or curve.

R^2 is a statistic that will give some information about the goodness of fit of a model. It provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model. The coefficient of determination ranges from 0 to 1 in most situations you will encounter.

- An R^2 close to one represents a trendline that is almost identical to the data points, meaning that you can use the trendline to accurately predict additional values. The closer the R^2 is to one, the better t to the data, an R^2 of exactly 1 indicates that the regression curve perfectly fits the data.
- You need several decimals accuracy of R^2 (usually 5 is good) for data that is very accurate, since there is a visual difference in the quality of t between R^2 values of 0.991 and 0.9993 data fits.
- A fit of 0.9 might sound pretty good, but actually the fit to the data can be rather poor.
- Often in life sciences and in data sets with several different confounding variables, R^2 values might be very low, like 0.5 0.7, yet they are considered to be indicative of a trend.

For more information on this subject consult this article and the references therein (Wikipedia on R Squared)

Question 4:

Let’s replicate a random error data point. Change the value in cell C9 to 1 and observe the changes in your trendline label.

What is your new R squared value, and what does that new value tell you about the accuracy of the (now changed) data?

Part II: “Advanced” Data Analysis in Excel

Purpose:

The remainder of this lab involves taking a brief look at what we’ll do in later labs this semester. For here you’ll manipulate data sets that we’ve taken in a more controlled environment to see what you’ll do once you take your own data.

Note

- There are tabs for every lab in this class that involves calculations, the goal is to get you to be comfortable using excel for all repetitive calculations, you should mostly not be using a calculator in this class.
- Some lab have blue text boxes for tips/explanations to help you. Be sure to follow them in order if they are numbered.

Exercise 1: Lab 4 - Inverse Square Tab

In this exercise you have measured distance and frequencies.

The data should follow what’s known as the optical inverse square law which says

$$I = \frac{C}{r^2}$$

For a constant C related to the source of the light.

If you take the log of both sides we can turn this into a linear form

$$\log I = -2 \log r + \log C \quad (4)$$

- A.)** Notice there are four graphs for this lab demonstrating two ways of linearizing the inverse square law.
- a. The first graph should currently look like an exponential, edit the axes to be “Logarithmic Scale” so that the graph appears linear. Note the type of trendline
 - b. The second graph (to the right) uses an equation to linearize the data. Observe the type of trendline used.

Question 5:

Compare equation (4) to the equation of a line.

- a. What should the value of the slope be?
 - b. How could you find the y intercept?
 - c. Does the sample data provided reflect the inverse square law? Support your answer using the appropriate trendline.
- B.)** Before leaving this tab delete the data given in the “Intensity Frequency” column, you’ll insert your own data here later.

Exercise 2: Lab 6 – Two Polarizer Tab

For this lab you will measure Angles and Intensities. You will change our system to a certain angle and measure the intensity for that angle. When doing this experiment you will do a full analysis of this data using excel, but for today you’ll just do the final steps.

Often in experimental work simple trendlines aren’t good enough and we have to do them ourselves for good data analysis. This experiment will utilize the solver to do this.

- A.)** Switch to the Two Polarizer tab, you should see two graphs.
- a. The top graph is a raw plot of the data, the red line is a generated fit.
 - b. The bottom graph is a linearization that is used for fitting.
 - c. When the data is properly fit the data points should be directly on the smooth lines.

Question 6:

Change the value you see for delta. Give a brief description of how changing delta affects the two graphs.

- B.) Read through the boxes in the sheet to get a good fit.
 - a. You'll repeat this again with your data when you perform this experiment.
- C.) Before leaving this tab delete the data in the Experimental Data "I(kHz)" column, you'll insert your own data here later.
- D.) Navigate to the Three Polarizer tab and repeat B and C.

Exercise 3: Lab 12 – Radiation 1 Tab

- A.) Navigate to the Radiation 1 Tab
- B.) This data set is another scenario where excel's built in trendlines aren't good enough to fit what we're working on.
 - a. The red smooth line is the trendline YOU are generating when running the solver to minimize the sum of squares here.
- C.) Follow the guide boxes in the page to use the solver to fit the data given.
 - a. Note you may have to run the solver twice to get a good fit.
- D.) Note the top graph shows the before, the bottom graph shows your fit.

Question 7:

Fill in the blanks of this trendline for your fit of the data given.

$$R = \text{ ______ } \exp(-\text{ ______ } * t) + \text{ ______ }$$

- E.) Before leaving this tab erase the data in the Experimental Data "Radiation" column, you'll insert your own data here later.

What You Need to Turn In:

You should have multiple graphs and some answered questions. Your instructor will decide what you need to turn in for this lab.

Remember you need this excel sheet the rest of the semester, SAVE IT SOMEWHERE SAFE (like email)