

Experiment 1

Introduction to 191 Lab

1. Introduction

In Physics 191 we will make extensive use of *Kaleidagraph* [*Kgraph*], a software package for graphing and data analysis, and *Excel*, for calculations. *Kgraph* is a substantial program but is nonetheless intuitive and extremely easy to use. Two major things that *Kgraph* does for us are:

1. Plotting high quality, easily customized graphs of your data.
2. Curve fitting, i.e., finding the parameters of a theoretical function that best describe your data. *Kgraph* provides estimates of the **statistical uncertainties of the fit parameters**, which are difficult to obtain with other commercial software (such as *Excel*).

2. Goals:

1. Familiarize yourself with basic calculations in *Excel*
2. Familiarize yourself with *Kaleidagraph's* plotting, calculation, and fitting options
3. Get a first impression of the style of labs and analysis for the class
4. Have a first look at histograms and standard deviation
5. Have a first look at fitting and residuals
6. Have a first try at searching for flaws (“systematic errors”) in data.

3. Background

3.1 In this first lab, we will learn how to use *Excel* and *Kgraph* on a PC equipped with the Windows XP operating system. Your lab report will contain printouts of the graphs that you generate. **The Reference Guide (RG), and your notes made while performing this lab, will also serve as your guide for doing basic analysis operations.** You may write your report outside of class by hand or with Word. But you must **use your time carefully** in class to plot and analyze your data. Since you will be changing lab partners regularly, **take turns, and be sure each of you understands the software.** *KEEP this write-up for reference in future labs.*

3.2 For this laboratory exercise, you will analyze information from an actual mileage log kept for a Prius automobile. The log usually contains an entry for each tank of gas purchased (missing data were occasionally estimated). The Prius uses a hybrid engine system, with both a gasoline engine and electric motors. The gasoline engine ultimately provides all the power, but a large battery stores energy when the engine is not heavily loaded, or when the car is braking. The electric motor uses energy stored in the battery to provide extra power after a stop, when passing, when going in reverse, or even to allow the engine to be turned off temporarily. The battery and electric motor allows the gas engine to run nearer its optimum efficiency, part of why the Prius gets more miles per gallon than most cars.

The odometer measures the distance (in miles) a car has traveled by counting the number of turns of the tires. The amount of gas purchased is calculated by the gas pump. The Prius has a dashboard meter that displays the miles per gallon (mpg) being obtained at the moment, and since the last reset (in this mileage log, that usually means since the last tank of gas). These readouts use the same distance measurement as the odometer, but the gas consumption estimate is based on the time (measured in microseconds) that the computer has opened the fuel injectors. This assumes that the fuel pressure is constant and therefore the amount of fuel consumed is proportional to the time the injector was open.

4 Excel Data Calculations Goal: print 1st page of modified spreadsheet

4.1 Use the desktop shortcut to the R drive to open R:\Computing_and_Graphing\Prius.xls, an *Excel* file containing the mileage log and some date information. **Now SAVE** to your computer's **U drive** space: File|Save As|U: (in Filename box)|Enter; pick your *section* folder and the *expl* folder inside it. An alternative to typing U: is clicking the down arrow in the Save in box and selecting bps1263xxxx. The Days column calculates number of days since the last fill-up. The *month* column calculates the months, starting from the beginning of 2002, so **month 7.5 is the middle of July, 2002**, and month 20 is the beginning of August, 2003. The Month of year column gives the month, independent of year, so month 9.3 would be early September of 2002, or 2003, or 2004. Click on a cell and the formula used will be displayed.

4.2 Now you will calculate some other quantities from the original data using *Excel* formulas. See the Reference Guide chapter on *Excel* if you need help. Calculate Miles (the difference in OD readings from the previous date to the current; use 456 for the miles on the first tank of gas); MPG Calc = miles / gallons; and calc/read = MPG Calc/mpg readout. Print the 1st page (File|Print|**Print Page(s) From:1 To:1**). Then **save and close your file**: *Kgraph* won't open the file if *Excel* still has it open.

5. Kaleidagraph: Histograms and Statistics Goal: print 2 histograms and record statistics

5.1 Navigating *Kgraph*

5.1.1. See the Reference Guide for a terse introduction to basic *Kgraph*; use as needed.

5.1.2. You can modify an existing graph by double-clicking or right-clicking objects to bring up a menu for that object: for example, the title of a graph, or an axis label.

5.1.3 *Kgraph* has a good Help system; the >> spins you through successive related entries.

5.1.4. If you want to print out the "Statistics" screen, or anything else you can't figure out how to print from a Windows program, hold down ALT and hit the Print Screen key. A copy of the active window is now on the clipboard. Then open Word, and paste with Ctrl-V, and print the Word document. The statistics screen or other clipboard contents can also be pasted into a *Kgraph* Plot, or a *Kgraph* Layout window.

5.2 Begin your analysis of the mileage data by opening *Kgraph*, and using *Kgraph* to open the Prius.xls file you just created. *Kgraph* properly opens an *Excel* spreadsheet with column headings as long as the headings are in the first row.

5.3 Make a histogram plot of the MPG Calc column. The plot will consist of bars rising from the x axis. In a histogram, data points are considered as being grouped into "bins" (ranges of values, say x between 20 and 30, and so on). The extent of the bar along the x axis will represent the

range of values of the variable in each “bin”, and the height of the bar (along the y axis) will represent how many data points fell into each bin. For your first plot, make the histogram with a bin size of 5 miles per gallon. You can find more about histograms by looking up the term in the index of Taylor.

5.4 For every plot in this entire class you should: Give the plot a title which explains the reason for this plot, such as Temperature Dependence of Velocity (metric units); no two plots should have the same title. Include your name, section, experiment, and date in the upper right hand corner of the graph (Help|Find|Text Tool). You can later copy/paste this text to other graphs. Make sure that the graph labels, legends, etc. do not overwrite any significant part of the graph. Label the axes of every graph with names and units, e.g., for a histogram Number of Log Entries and Calculated MPG (mi/gal); or for a scatter plot, say Velocity (m/s) and Time (months). **Print your histogram and save it.**

5.5 Now make and print the histogram with a bin size of 2 miles per gallon. Does the new histogram look the same as the previous graph? Why or why not? Did you learn anything new by changing the bin size? **As you write in your lab notebook, or lab report, always include the number of the part of the write-up (e.g. 5.5).**

5.6 Now you will use one of the advanced functions of Kgraph to analyze the data. Click on the frame of the data sheet to highlight it; then click on the MPG Calc column, and then click Function|**Statistics** as described in the Reference Guide. Note the entries for Mean and Std Deviation. The Mean is just the average value; Std Deviation is an abbreviation for Standard Deviation, a measure of the scatter of the values around the average. Both are discussed in Taylor. Write these two values down in your notebook or paste the statistics block onto one of your histograms. On the histogram with 5 mpg bins, indicate the location of the mean, and draw a horizontal line stretching from (mean – standard deviation) to (mean + standard deviation).

6. Kaleidagraph: Scatter Plots and Fits Goal: print mpg plot, fuzzy parabola fits and residuals.

6.1 **Scatter Plot** Now you will seek a better understanding of the mpg data by plotting it as a function of time. Make a scatter plot of calculated mpg as a function of *month*, that is, mpg vs. time. **Whenever we say “y vs. x” we mean the first variable should be on the y (vertical) axis** and the second should be on the x (horizontal) axis. Do not use the Line Plot, as it is a “connect the dots” plot not well suited to our uses: by default it doesn’t display much of the data, and the lines both over-lead the eye and get in the way of fits we will apply to the data. Make the data 18-point hollow diamonds (e.g. Plot|Variable Settings|Marker size|18). Label the axes of the graph with names and units, e.g., Velocity (m/s) and Temperature (K). Does this graph tell you anything about the reason for the structure you saw in the mpg histogram? How would you explain it? *Hint*: look at the definition of *month* in section 4.1.

6.2 Next you will “fit” a mathematical model. The “fitting” consists of writing down a mathematical expression (“curve”) with some unknown parameters and letting Kgraph automatically try values of the parameters to find those that best “fit” the data. The mathematical method Kgraph uses to find the parameters is called “Least Squares”. Least squares fitting minimizes (finds the *Least*) the sum of the *Square* of the difference (“residual”) between the measured data and the values predicted by the curve. This will be discussed in detail in Chapter 8 of Taylor.

To get a feel for fitting, you will now work through an example where the model is perfectly known, and see how well *Kgraph* reconstructs the true model. Open and save to your U: drive the file `r:\exp1_ref\fuzzy_parabola.xls`. Then close it and re-open the *copy* in *Kgraph*. Make a scatter plot of $y = f(x)$ vs. x . Now fit the model $a + bx + cx^2$ to the data using the directions in the Reference Guide. Write down the values you obtain for the parameters a , b and c . The curve should pass perfectly through the data and *Kgraph* should report essentially zero uncertainty (“error”) in the parameters. Check that you got the correct values by opening the `fuzzy_parabola.xls` file with *Excel*. Look at the equation used for the $f(x)$ column and write down the correct values of the parameters in your notebook (and lab report).

6.3 Now make a new plot of *fuzzy f* vs. x . The “fuzzy” value was created by adding a simulated “measurement error” to the value of f . (It is impossible to make a measurement which returns the exact “true” value). I used a random number between -1 and $+1$; the last column gives the difference between the fuzzy and real value. Fit the *fuzzy f* data to the parabola. Print the plot with the fit, the values of the parameters and their uncertainties. How much do they differ from the correct values?

6.4 Now look at how the fit differs from the data by examining the “residuals”, defined as the difference: $\text{residual} = \text{data} - \text{fit}$, at each data point. *Kgraph* uses $\text{fit} - \text{data}$, but that isn’t a big problem. Make and print a histogram of the residuals (see Reference Guide). You should see values roughly between -1 and 1 : the differences between the fit and the data are about the same as the “fuzz” we put in. If we’d looked at the residuals of the fit to the un-fuzzed $f(x)$, the values would have been tiny.

7. Data Analysis. Goal: print fit to mpg vs. time; % difference between calc and readout

7.1 Now return to the mpg vs. month data. There isn’t a precise theory of how mileage should vary with time, so the model will be approximate.

7.2 Make a fit with the **cosine** curve as described in the Reference Guide. You will need reasonable starting values for this fit to work properly. Print the plot with the fit curve and the fit parameters as labels for the figure. Also, histogram the residuals from this fit. Use the statistics function to calculate the standard deviation of the residuals: this, as you just saw, is a measure of how well the fit matches the data.

7.3 Now **cross check** the two estimates of fuel consumption to see if they’re roughly equivalent, and how close they are to each other. You can approach this question with the techniques you’ve learned: plotting or examining statistics for the column mpg calc/readout, or making a scatter plot of the calculated vs. readout miles per gallon and performing a fit. Choose one of these techniques and find quantitatively, by how many % the two estimates differ typically, and how much (in %) individual values vary about the main trend.

7.5 Finishing up: As you exit *Kaleidagraph*, select “None” in the dialogue box when you are asked whether you want to save changes. **It is advisable to back up files from the U Drive to a USB flash drive** if you will need your data for next week (**the U drive is not backed up!**):

- 1) Insert USB flash drive in front of the computer, or on USB extender
- 2) Open the U drive by double-clicking on the shortcut
- 3) Select in the Address box of this window your flash drive (F: or G:)

- 4) Again open the U drive by double-clicking on the shortcut (gives a second window)
- 5) Drag the files or folders from the U: window to the flash drive window
- 6) **Left**-click Safely Remove Hardware icon in system tray (bottom right)
- 7) Click Remove F; after it says it's OK to do so, actually remove your flash drive.
- 8) Next time: Pick your program | Open |Type F: in "filename" box and hit Enter

You may also email the files to yourself.

8. Summary Tables In each lab report you should present your critical results in summary tables which organize your results for you, and for your readers. In many experiments you will also need to organize your data recording in tables. At the beginning, we will give you explicit tables to use; as time goes on, you should be able to think through for yourself what tables you will need. For this experiment, your summary tables should look like the following:

| Quantity | Mean | Standard Deviation |
|----------------------------------|------|--------------------|
| MPG Calculated | | |
| MPG Cosine Fit Residuals | | |
| MPG Calculated / MPG Readout (%) | | |

| f(x) fits | a | b | C |
|-------------------------------------|---|---|---|
| true parabola parameters | | | |
| f(x) fit parameters | | | |
| fuzzy f(x) fit parameters | | | |
| uncertainty of fuzzy fit parameters | | | |

To finish the report, read Taylor chapter 1, 2 (you may skip sections 2.6 and 2.9), 4.2, and 5.1 . Read section 1.3 and chapter 2 especially carefully—they give the *purpose of uncertainty calculations*, to be able to make *quantitative comparisons* between measured values and between measurements and predictions. See also the Reference Guide (section 4) on the important points on uncertainties.

If you have trouble using excel, complete the **Excel Tutorial** in the Reference Guide (section 2).

9. Questions to be discussed in your report:

- 9.1 Why should one look at histograms with different bin widths?
- 9.2 What do you think causes the structure you see in the finer-binned mpg histogram?
- 9.3 Are the parameter "errors" (the uncertainties given *Kgraph*) in the parabola fits about the size of the difference between the "fuzzed" parameters and the original ones? Is this what you would expect, roughly?

10. Searching for Systematic Errors (Incompleteness of Models): Where could the measurements be misleading? Any experimental procedure is subject to errors of execution, mistaken assumptions, or biases. Part of our job as experimentalists is to think about them. At the beginning of the term, we will guide your thinking about these, but you should strive to be

able to do this more independently as the term progresses. Here we will suggest some sources of uncertainty in the data and procedures.

10.1 What would happen to the miles per gallon if the tires were under-inflated? By about how many percent might under-inflation affect the estimated distance traveled? Would under-inflation cause a measurement to be too low, or too high?

10.2 The calculated mpg depends on the precision of the measurement of gasoline volume by the gas station pump. What do you estimate for the fractional % uncertainty in the number gallons?

10.3 Is the scatter in the data about the fits larger or smaller than your estimate from 5.6? If larger, what other effects might be causing such variations from one tank to the next?

10.4 Is the mpg according to the readout systematically different from the calculated value? By how many % If so, which measurement would you tend to believe more? Why?

Hints for section 10:

10.1 Speedometers are often off by 2 - 5 mph at 65 mph, so a 3 - 8% error in the speedometer is not out of the question. However, that's the speedometer, which is perhaps more complicated than the odometer. The odometer works by counting the revolutions of the tires and converting turns into distance traveled. This assumes the tires are of constant circumference. The accuracy implied by our uncertainty was $\approx 1\text{mi}/456\text{mi} = 0.2\%$ which seems awfully strict when you think of the change of shape in the tires when they are fully inflated vs. under-inflated. You might estimate the tire height (diameter) might change 1/2 inch out of about 2 feet, or about 2 %, due to changes in tire pressure, from being warm or cold, or over or under-inflated. (This alone might give a 1 mpg apparent decrease for highway driving, which keeps the tires warm and records fewer apparent miles (axle revolutions) than actual miles, giving a lowered mpg).

10.2 For a first estimate use the recording error of .1 gal; for 11.3 gal that's 0.9% (though the pump records more decimal places, that's all that was recorded in the log). Assume that all pumps are calibrated by the state to measure the same within well less than the 0.9 % measurement error on a tank (at least we hope so). What about the definition of a full tank? Is that repeatable from pump to pump? What is its uncertainty? Hard to guess again but it might be another 0.1 gallon, or perhaps even more. This latter might not matter so much in the long run: if the pump cut off early for one tank, you'd miss some of the gas you actually used for those miles. But the next tank, you'd probably record some gas for miles which really were recorded for the previous tank. Over time it would average out, but this uncertainty would contribute to the uncertainty of any single tank.

10.3 If the scatter of measurements is large compared to your estimated error (as seen by either the calculated vs. readout data, or the size of the fit residuals), then the measurement was dominated by sources of uncertainty other than those you have thought about. These are sometimes called uncontrolled variables. They might not have anything to do with the measurement process itself, but rather changes in what is actually being measured. A driver might be more interested in speed than good mileage on a particular trip. Weather, traffic, road conditions, or the load carried might vary.

