Experiment 1

Statistics and Uncertainty

1.1 INTRODUCTION

The local acceleration of gravity, g, is an important parameter for many aspects of geophysics and planetary science. Slight variations in g across the surface of a planet provide clues about underlying geological features, such as the existence of mineral deposits or the presence of subsurface water. For such measurements to be useful, however, it is important to have a good understanding of the uncertainties involved.

In this experiment, you will use a simple pendulum to estimate the value of g. More importantly, you will learn how to compute a mean value and its uncertainty from a series of repeated measurements, how to propagate uncertainties, and how to compare experimental results with expected values.

The write-up for this experiment is longer and more detailed than most of the experiments in this course because you will be using ideas from this experiment in almost every other experiment this semester. Accordingly, you should work carefully and ensure you understand every step of your work. Include any useful notes to yourself right in your lab writeup, so that you can refer to them later in the semester.

1.2 THEORY

An ideal simple pendulum consists of a point mass m at the end of a massless string of length L. For small oscillations, the period T of a pendulum (*i.e.* the time it takes to complete one back-and-forth swing) is related to its length L and the local acceleration due to gravity, g, by

$$T = 2\pi \sqrt{\frac{L}{g}} \; . \tag{1.1}$$

Consequently, if you measure L and T, you can determine g:

$$g = 4\pi^2 \frac{L}{T^2} \; . \tag{1.2}$$

Note that the mass m doesn't matter. This is the main equation that you will use today.

1.3 GETTING STARTED

All work will be submitted online as a single PDF document via a Moodle Assignment. These instructions will assume you are using Microsoft Word and Excel on the lab computers, but you may also use Google Docs and Sheets, or any other tools you find convenient.

Start by creating a new document. At the top, place your name, your lab partner's name, the date, and the name of the experiment. Give your file a sensible name, such as `Phys131-Lab01-yourNetworkId`. (For Microsoft Word, use the `File→Save As` menu item. For Google Docs, use `File→Rename`.)

In most experiments, you will have to perform a number of calculations. Although you can do them all by hand and then type the results into your report, it is often better (more efficient, less error-prone, and easier to read) to do the calculations in a spreadsheet and then include that in your report.

Create a new Excel spreadsheet. You will later be able to link (or copy and paste) this sheet into your report. At the top, in cell A1, enter the lab section and lab title. In cell A2, enter your name and your lab partner's name. In cell A3, enter today's date. Save your work.

1.4 EXPERIMENT

Ideally, a simple pendulum consists of a point mass at the end of a massless string. In practice, a length of string roughly 1 meter long with a small weight attached to the end will do fine.

You will also need a timer. A stopwatch is available at each lab table, Most smartphones also have a suitable stopwatch app. If your stopwatch app has a "lap timer," that can be very handy for doing multiple trials.

1.4.1 Data Acquisition

Measure the length L of your pendulum, from the attachment point to the center of the mass at the end of your pendulum. (You can estimate the center by taking the average of the distance to the top and the distance to the bottom.) It is important to measure this distance well, so take your time and do this step carefully. Record the length in your lab report, and express the result in meters.

Make an estimate of the *uncertainty* in L. This is not a statistical estimate, it is just a judgment you make based on how you took the measurement. There are a number of minor complications to consider: You have to estimate numbers off a ruler. The string stretches a little. You have to measure to the middle of the pendulum. Do you believe you can measure the length to an accuracy of $\pm 10\,\text{mm}$? $\pm 1\,\text{mm}$? There is no "correct" answer here, just a reasonable estimate of how certain you are of your best estimate of L. Record this uncertainty in your lab report.

Enter this data into your spreadsheet. In cell C6, put a label 'L (m)'. In cell D6, put the length of your pendulum, in meters. In cell E6, put your estimate of the uncertainty in L, again in meters. Add two more labels: In cell D5 put a label 'Value' and in cell E5, put a label 'Uncertainty'.

Start the pendulum swinging gently with a relatively small amplitude (a few centimeters back and forth). Let it swing back and forth several times and then use your stopwatch or timer to record the time it takes to complete 5 full back-and-forth swings. Don't try to release the pendulum and start the timer at the same time. Instead, release the pendulum, watch it swing back and forth a few times, and then start timing at the beginning of one swing.

Repeat 4 times (for a total of 5 trials). If your timer app has a "lap timer" feature, you may be able to count 5 complete swings as one "lap," and then time a sequence of 5 laps. Switch places and have your lab parter time an additional 5 trials.

You should enter these times directly into Excel. It is important to give columns clear labels, with units. In cell A5, enter a title such as '5 swings (s)', for your times. You will enter your 10 times in cells A6 through A15.

Look carefully at your 10 times. They should all be about the same. (We will be more precise about that in a moment.) If you notice any problems or irregularities in your data, try taking a fresh set of timings. If you are unsure, consult with your instructor.

Before continuing with the analysis, go back and check the length. Do not disassemble the pendulum—you may need to come back and recheck your results.

1.4.2   Calculating the Period

Since you measured the time for 5 complete swings, the period is actually one fifth of that time. Create a new column for the period T. In cell B5, put the title 'T (s)'. In cell B6, put the formula

$$=A6/5$$

Note that formulas in `Google Sheets` and `Excel` begin with an equals sign "=".

Copy and paste cell B6 to cells B7 through B15. Click in any of those cells to see how the formulas automatically adjust to point to the correct cells immediately to the left.

1.5   DATA ANALYSIS

1.5.1   Systematic Uncertainty and Random Fluctuations

Physics, like most science, is quantitative in nature. An experiment is typically designed to produce a numerical result. Experimental results might be compared to values predicted by theory, or to standard values—for instance a test of whether a system being studied meets a customer's design specifications.

To make such comparisons, we need to consider a general question: given an experimental result and a value to which it is to be compared, how do we decide whether or the difference between them is significant? Have we discovered something important, or is the difference simply due to *measurement uncertainty*? This concept is important in all fields of science and engineering, and we will make use of it in almost every lab we perform this year. This concept is also closely related to concepts such as tolerance and precision in manufacturing processes.

All measurements contain at least some uncertainty. It is useful to distinguish between two types of uncertainties: systematic and random.

Systematic Uncertainty

If you measure a length with a meter stick that has shrunk, your results will contain *systematic uncertainty*. That is, your results will always deviate from the true value in the same direction. If you had some way to determine the factor by which the meter stick had shrunk, you could correct all your data accordingly. However, systematic error is rarely easy to quantify. After making your best efforts to minimize it, you continue the experiment.

Similarly, if you are timing an event with a stopwatch, and always start the watch too early or too late, then your measured times will be *systematically* in error. Fortunately you can often minimize systematic uncertainty by using different people to do the timing and by

designing the experiment so that the typical human reaction time is a small fraction of the total time measured.

Random Uncertainty

If you repeat the same experiment many times, and if you measure the results precisely enough, you will always find small, random variations in the measurements. This is called *random uncertainty* and results from the inevitable imprecision of the experimental apparatus available to you. Random uncertainty also arises from the quantum nature of matter, which is *inherently uncertain* in its behavior, but that is a subject far afield from the present class.

Sometimes, you can reduce random uncertainty by increasing the precision of your equipment (e.g., using better rulers or timers) and by performing the experiment with greater care. However, it is impossible to eliminate it completely. You will always find *random scatter* in your measurements. If you repeat a measurement many times and average the results, these differences will tend to cancel and your result will be closer to the true value. The results still contain random uncertainty, but the averaging process tends to make your results more reliable. Your task is to determine that average value *and its uncertainty*.

Application to this Experiment

In this experiment, you will determine g by measuring T and L, and using Eq. 1.2. Since you made multiple measurements of T, those values will show *random* uncertainty. Some are likely too high, while others are likely too low. By taking multiple measurements, we hope those random fluctuations will tend to balance out. On the other hand, you only have one value for L. If your value for L is too high, then your estimate for g will also be too high. That is an example of a *systematic* uncertainty.

1.5.2 Means and standard deviations

Mean, $\overline{T}$

Ideally, all of your times T measured above would be identical. In reality, however, there is likely to be some scatter—some values will be too high while others will be too low. If these variations are random, then the best estimate of the "true" value is the *mean*, given by

$$\overline{T} = \frac{1}{N} \sum_{i=1}^{N} T_i \tag{1.3}$$

where N is the total number of trials (10, in this experiment) and T_i are your individual period measurements in column B.

You shouldn't calculate Eq. 1.3 directly—most calculators have a special key to calculate the mean for you; often it is indicated by $\bar{x}$. In `Google Sheets` or `Excel`, you can use the `AVERAGE()` function to calculate the mean.

In cell A17, enter the label 'Mean'. In cell B17, enter the formula for the mean:

$$\text{=Average(B6:B15)}$$

Record this mean time $\overline{T}$ in your report.

Standard Deviation, σ

Now that you have found the mean value, it is also important to characterize how much your measured values scatter around that mean. The most commonly used measure of that scatter is the *standard deviation*, given mathematically by

$$\sigma = \left[\frac{1}{N-1} \sum_{i=1}^{N} (T_i - \overline{T})^2 \right]^{1/2} . \tag{1.4}$$

Again, you shouldn't calculate Eq. 1.4 directly—most scientific calculators have a key for this also; some have two keys, one labeled σ_n and another labeled σ_{n-1}. Make sure you use the latter: a key labeled σ_n most likely calculates a similar quantity, but with just N in the denominator, not $N-1$. Consult your calculator manual or refer to the example in Appendix A if you have any doubt. In `Google Sheets` or `Excel`, you can use the `STDEV()` function to calculate the standard deviation. Specifically, in cell B18, enter the formula

$$\text{=STDEV(B6:B15)}$$

Again, labels are helpful, so in cell A18, enter the label 'Std. Dev.'

The standard deviation tells us, in effect, how far from the average the next single measurement is likely to be. More specifically, if your data is "normal" in some sense, then you should expect about 68% of your trials to give results in the range $(\overline{T} - \sigma)$ to $(\overline{T} + \sigma)$; about 95% should fall within 2σ of the mean, and 99.7% should fall within 3σ. These limits are usually called "one-sigma," "two-sigma" and "three-sigma".

Calculate the standard deviation for your period measurements and record it in your report.

Go back and look at your data table. Are most of your values within 1σ of the mean? Are any more than 3σ away from the mean? If any points seem inconsistent with the rest, you may go back and get more timing data. If in doubt, consult with your instructor. It is much easier to replace a bad value now than to wait until the end when you might have to redo all the calculations. If you make any changes, document them in your report.

Uncertainty of the Mean, $\sigma_{\overline{T}}$

Lastly, you should estimate how *reliable* your mean value is. If you took another set of ten trials and averaged those results, how different might the new average be? One measure of this reliability (sometimes called the *standard deviation of the mean*), is given by

$$\sigma_{\overline{T}} = \frac{\sigma}{\sqrt{N}}. \tag{1.5}$$

We will call this the *uncertainty* of the measurement $\overline{T}$.

In `Google Sheets` or `Excel`, you can calculate this by entering the following formula in cell B19:

```
=B18/Sqrt(Count(B6:B15))
```

The `Count` function counts the number of entries in the given range, which is 10 in this case. The formula then takes the standard deviation (from cell B18) and divides it by $\sqrt{10}$. Label this by putting 'Uncertainty' in cell A19.

The distinction between the *standard deviation*, σ, and the *uncertainty*, $\sigma_{\overline{T}}$, is an important one. If you ask how far from the mean the next single trial is likely to be, it should not matter whether you have done ten trials or a hundred trials; the *standard deviation* should not change significantly when you increase the number of trials substantially. It just reflects the scatter of the individual trials.

On the other hand, the more trials you perform, the more reliable you should expect the *average* result to be. So the *random uncertainty* in the mean value should get smaller and smaller as you increase the number of trials. The $1/\sqrt{N}$ factor in $\sigma_{\overline{T}}$ does just that.

Record the uncertainty of the mean $\sigma_{\overline{T}}$ for your time measurements in your report.

Reporting Results

When you report the results of an experiment, it is very important that you also report a measure of their reliability. Therefore you usually should report mean values *with their uncertainty*: $\overline{T} \pm \sigma_{\overline{T}}$, not $\overline{T}$ alone. For example, if you calculate a mean of 3.59284 seconds for your data, and the uncertainty of the mean turns out to be 0.629182 seconds, you would report the result as 3.6 ± 0.6 seconds. Notice that the uncertainty of the mean is typically rounded to only a single significant digit, and the mean is rounded to match (*i.e.*, at the tenths place since that is the point at which $\sigma_{\overline{t}}$ is truncated). It makes no sense to retain any digits beyond these; they are meaningless. Also note that the mean value and the uncertainty always have the same units.

Record your best estimate of the time it takes for the pendulum to complete one full swing, $\overline{T}$, and the uncertainty $\sigma_{\overline{T}}$ in that time.

It is helpful to record these in the spreadsheet as well: In cell C7, put the label 'T (s)'. (This is the row under your L measurements.) In cell D7, put the formula "=B17" to copy the mean value for the period, and in cell E7, put the formula "=B19" to copy the uncertainty.

1.5.3   Results

At this point, you have calculated the mean time $\overline{T}$ it takes a pendulum of length L to complete one back-and-forth swing. But the goal of the experiment is to find the gravitational acceleration g. How do you find your best estimate for g and the uncertainty in that measurement?

We will study oscillations later this semester, but for now, simply use Eq. 1.2 from the theory section:

$$g = 4\pi^2 \frac{L}{T^2}$$

Use your mean value of $\overline{T}$ (from Eq. 1.3) and your value for L in Eq. 1.2 to calculate your best estimate for g. You can do that right in `Excel` with a formula in cell D8:

```
=4*PI()^2*D6/D7^2
```

Note that you enter π by typing the function `PI()`, and you square a number by entering `^2`. You had previously entered the length L in cell D6 and the mean period in cell D7.

As you may know, the expected value is $g = 9.81 m/s^2$. Your value should be close to this, but likely differs by a small amount. How can you tell if that difference is significant? To answer that question, you need to consider the uncertainty in your measurements.

Propagation of Uncertainty

Next, you need to calculate the uncertainty in your best estimate of g. Clearly it is related to the uncertainties in the quantities L and T that go into the calculation of g, but how do you calculate it?

In general, the process of determining the uncertainty in a final result based on the uncertainties of the component quantities is known as Propagation of Uncertainty. The detailed rules for propagation of uncertainty are described in Appendix B, distributed with the lab overview material. Here, we will focus on the specific case at hand.

In this experiment, both L and T have uncertainty, but the time uncertainty is typically more significant. Accordingly, we will for the moment ignore the uncertainty in L. We will revisit that assumption later.

It will help to develop some notation. Let $\overline{T}$ be your best estimate of the period (from Eq. 1.3), and let $\sigma_{\overline{T}}$ be the uncertainty in T (as calculated by Eq. 1.5). Also, let g be your best estimate of the gravitational acceleration (from Eq. 1.2). Your goal here is to estimate $\sigma_{\bar{g}}$, the uncertainty in g.

Equation 1.2 shows that g is proportional to T^{-2}. Appendix B gives a rule for this circumstance. The uncertainty of g (which we want to know) is related to the uncertainty of T (which we have already calculated) by the following ratio rule:

$$\frac{\sigma_{\bar{g}}}{g} = 2\frac{\sigma_{\overline{T}}}{T}. \tag{1.6}$$

The factor of 2 reflects the fact that T is squared in Eq. 1.2. Equation 1.6 can then be rearranged to give $\sigma_{\bar{g}}$:

$$\sigma_{\bar{g}} = g\left(2\frac{\sigma_{\overline{T}}}{T}\right) \tag{1.7}$$

Calculation in Excel   To calculate this in Excel, place a label 'Fract. Unc.' in cell F5 for the fractional uncertainties. In cell F7, use the formula

=E7/D7

to calculate $\dfrac{\sigma_{\overline{T}}}{T}$, the fractional uncertainty in T.

To calculate Eq. 1.6, the fractional uncertainty in g, place the following formula in cell F8:

=2*F7

Finally, to calculate the uncertainty from Eq. 1.7, place the following formula in cell E8:

=D8*F8

Save your work.

Uncertainty in g   Use Eq. 1.7 to determine the uncertainty in your estimate of g. The result should already be in cell E8 of your spreadsheet. Report your finding as $g \pm \sigma_{\bar{g}}$. Be sure to round off your results appropriately. (That is, round $\sigma_{\bar{g}}$ to one significant figure, and then round $\bar{g}$ off to the matching decimal place.)

Back to your Report   Now is a good time to insert your spreadsheet into your report document. In Microsoft Word, Select the **Insert** menu. In the **Text** panel, select the unlabeled **Object** icon from the lower right, and then select **Object**.
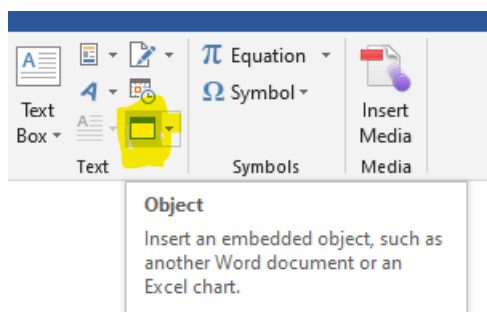
Select the `Create from file` tab and browse to insert your Excel spreadsheet. Click the `Link to file` box so that changes to your spreadsheet are automatically reflected in your Word document. (If the whole spreadsheet is not shown, try saving your Word document, quitting, and then re-opening your Word document.)

If you are using Google Sheets, highlight the useful cells in your sheet, and use the menu item `Edit`→`Copy`. Then, in your `Google Doc` report, select `Edit`→`Paste`. Select the option "Link to spreadsheet" and then select "Paste." This ensures that as you update the original spreadsheet, your report can easily be updated as well.

Comparison With Expected Values

Your estimate of g is probably slightly different than the expected value of 9.81 m/s^2. The next challenge is to determine if the difference between these values is significant.

What is the difference between your value of g and the standard value of $9.81 m/s^2$? How does that difference compare to the uncertainty $\sigma_{\bar{g}}$? If the measured and expected values differ by less than $\sigma_{\bar{g}}$, then we say the results are consistent. If the measured and expected values differ by more than $3\sigma_{\bar{g}}$, then we say the results disagree. If the difference is more than $\sigma_{\bar{g}}$ but less than $3\sigma_{\bar{g}}$, then the situation is unclear. Ideally, you would go back and take more data, since random errors will tend to cancel out. Practically, however, you often won't have time to go back and take more data, so you should just state your results clearly and then move on.

Do your results agree with the expected value to within the uncertainty?

Optional: Including the Length Uncertainty   If the uncertainty in length L is also important, then Eqs. 1.6 and 1.7 need to be modified. Specifically, since g is proportional to L, we also need to include the ratio σ_L/L. Since the uncertainties in T and L are unrelated, the standard practice is to combine them in the following way:

$$\frac{\sigma_{\bar{g}}}{g} = \sqrt{\left(2\frac{\sigma_{\bar{T}}}{T}\right)^2 + \left(\frac{\sigma_L}{L}\right)^2} \qquad (1.8)$$

Multiplying both sides by g gives the final equation for the uncertainty in g:

$$\sigma_{\bar{g}} = g \sqrt{\left(2\frac{\sigma_{\bar{T}}}{T}\right)^2 + \left(\frac{\sigma_L}{L}\right)^2} \tag{1.9}$$

Discussion of Sources of Uncertainty

If your results disagree, then you should try to figure out why. Most of the experiments you will do this semester should work, to within the experimental uncertainties. If you get results that disagree, you should try to figure out why, perhaps with the help of your instructor. Sometimes, there is a measurement error. Sometimes there is an error in a calculation. Sometimes there is an error in the theory used. It may take some persistence, but you should do your best to track down sources of disagreement and eliminate them. **If time permits, you should fix any mistakes**.

In your report, discuss sources of uncertainty. *A long list of possible errors is usually not very useful. Instead, try to identify the one or two most important problems.* Adopt a "try it and see" approach, if feasible.

For example, suppose you tended to stop the timer slightly before the pendulum completed the fifth swing. Would this cause an error in your estimate of g? Would this error tend to make your estimate too high or too low? Do you think this might have been a big effect or a small effect (where big and small are relative to your uncertainty)? Try it and see. Deliberately stop the timer a little bit too early. What happens? Be quantitative. Was this error likely to have been an important source of error in your experiment today?

Similarly, consider the length. We have assumed the uncertainty was negligible, but is that a fair assumption? How does the fractional uncertainty in length $(\frac{\sigma_L}{L})$ compare to the fractional uncertainty in time $(\frac{\sigma_{\bar{T}}}{T})$?

What happens if you go back and check the length again? (Note you can try plugging different numbers into cell D6 to see what happens.) Was this likely to have been an important source of error in your experiment today?

These are specific examples of the sort of inquiry you ought to try each week. When considering a possible source of error, you need to consider the size of the effect, the direction of the effect, and the likelihood of the effect actually occurring before you can adequately assess the importance of that source of uncertainty.

It is not always this easy to test a possible source of error, but you should try things out when you can.

Suggestions for Improvement

Finally, look back over your whole experiment. Are there any significant ways in which the experiment or write-up could be improved? We always welcome helpful suggestions.

1.6   FINAL REPORT SUBMISSION

Take a final look at your report to make sure it is complete. If you want to add anything (*e.g.* a sketch or photo of your apparatus, or a scan of any hand calculations) you may do so. Make sure you include the division of labor in your report.

Remember that your report is intended to convey your findings to the reader. It needn't be a highly polished document, but it should be clear and easy to follow.

Create a PDF of your document. Check the file on your local computer to make sure it is correct, and then upload it to the appropriate Moodle Assignment.