LAB 01 – Experimental Errors and Data Analysis

CONTENT:
1. Introduction
2. Experimental Issues
3. Experimental Errors
   A. Classification
   B. Calculations
   C. Reporting
4. Graphing Experimental Data
   D. Principle and Design
   E. Analysis
5. Pre-lab:
   A. Activities
   B. Preliminary info
   C. Quiz

1. Introduction

As we emphasized a few times in class, Physics is a natural science driven by scientific method, which is a standardized process designed to build a model of reality devoid of subjective interferences, such as cultural prejudices and biases. In its standard form, the scientific method employs three steps where a conclusion is reached based on a set of experimentally validated premises:

1. Observation focused on certain aspects of a natural phenomenon.
2. Formulation of a hypothesis deemed to explain the phenomenon tentatively via a descriptive or mathematical causal mechanism. To be acceptable as a scientific hypothesis, the putative mechanism must be falsifiable experimentally: that is, it must be capable to make testable predictions about the evolution of the respective phenomenon.
3. Experimental validation of the predictions in a reproducible manner by independent experimenters. Once the predictions are tested, the hypotheses can be elevated to the status of scientific theory acceptable within the limits of the experimental setup used to verify it. It is important to understand that science abhors hubris and it remains a healthy endeavor only as long as it is practiced with a creative dose of skepticism. It is a common occurrence in the history of science that mighty long-standing theories about natural phenomena to be eventually disproved or reduced to merely particular cases of more general models. For example, look no further than to Newtonian mechanics (about to be introduced as an archetypal science this semester) which dominated natural sciences for two centuries before it was dethroned to the status of a macroscopic low-speed limit of quantum mechanics and relativistic mechanics.

Note that there is nothing esoteric or elitist about the scientific method. You can apply it to otherwise prosaic situations such as fixing a problem with your car. The first epiphany in one’s quest to understanding science ought to be realizing the effective simplicity of its basic methods.

To frame properly the subject of our first PHYS-154 lab, and of this lab class in general, let us take a closer look at the experimental testing which reigns supreme in validating hypotheses. In particular, when processing experimental records, plotting data is often the most powerful method to visualize and infer relationships between quantities. Hence, in our lab class we shall systematically graph numerical results, and you ought to review some basic rules of building a neat, correct and complete graph.

2. Experimental Issues

As highlighted above, experiments cannot confirm anything ultimately. In fact, it is important to note that the purpose of the scientific method is not to prove hypotheses correct, but tentatively not false, because experimental results are valid only within a limited range of measurable parameters which can be used at most to define a domain of applicability rather than justify any pretensions of universality. Incidentally, this nuances the example above with Newtonian mechanics: we see that its birth, longevity and popularity is hardly fortuitous inasmuch as the domain of applicability of its models
overlaps with the domain of our everyday experience. Thence, in Physics we talk about the “classical limit” which sets the adequacy of Newtonian mechanics only for objects much larger and more massive than particles of atomic scale, and moving much slower than light in vacuum.

How come experiments are inherently limited to certain domains of relevance? And how do we actually measure and quantify the uncertainty associated with a particular experimental setup? In brief, experimental relevance is limited both by apparatus and the intrinsic nature of measured objects. Reality has “blurry edges”; it actually comes to appear fairly well defined only due to the macroscopic coherence arising statistically from an otherwise chaotic microscopic structure. To quantify and put in critical perspective this uncertainty of particular experiments, we use error analysis performed on experimental data.

Organizing and conducting an experiment can be a rather complex job, demanding that one applies ingenuity to very particular physical situations. Yet, many experiments follow a common strategy where a quantity – call it an independent variable – is changed whilst another quantity – call it a dependent variable – is monitored; meanwhile other factors – call them control variables – are kept constant. Thence, inasmuch as the two types of variables are related via the very mechanism that is being tested, the experiment will serves as a validation machine for the respective hypothesis.

Example 1: Imagine that you want to verify an idea that you will hear in this class: considering an object sliding on a slippery surface, in a first approximation the contact friction will depend only on the nature of the two surfaces in contact and the weight of the object. This is a somewhat disconcerting statement. For once, it implies that the friction is independent of how wide is the contact area. So, in a fit of critical thinking, you may demote the idea to the status of an unverified hypothesis, and perform an experiment to confirm it using a block of wood with faces of different areas. The block could be launched on a long table to see if the contact area has an impact on friction. In this case, the independent variable would be the face of the object, the dependent variable could be the stopping distance (a measure of the strength of friction), while a few other parameters would be control variables, such as the weight and the initial speed of the block. The hypothesis will be confirmed if the stopping distance remains the same irrespective of the sliding face.

Note that the relevance of an experiment may be affected by other factors rather than objective inherent errors. One such factor is dishonesty. Another is the unconscious or negligent cherry-picking of data to prop the preference or bias of the researcher for a certain hypothesis. However, this is not the kind of errors that we shall be studying in this lab. We shall graciously assume honesty and responsibility in our data collection, analysis and reporting. So let us focus on the unavoidable errors associated with experimental setups.

3. Experimental Errors

A. Classification

For completion and scientific respectability, the errors affecting an experiment must be specified. Not only that they are unavoidable, but they are technically relevant as part of the clockwork of data analysis. Additionally, in an instructional setup like ours where we merely test well established models, errors can be used to compare the performance of different lab teams, and to check the results against published data.

Experimental errors are typically classified into systematic (determinate) and random (indeterminate or statistical):

Systematic errors are typically due to faulty instruments or measuring techniques affecting the measurement in only one direction by the same deviation.

Random errors are due to the inherent variations in the measured characteristics due to uncontrollable causes such as thermal fluctuations, tiny vibrations, or microscopic irregularities. The impact of these erratic variations can be reduced, but never completely eliminated. Therefore, repeated measurements of the same quantity in the same setup will result in numbers that will be distributed on both sides about an average value within a range proportional to the amplitude of the error. Thus, the spread of the data about this average is a measure of the uncertainty of the measurement.

In lab parlance, the accuracy and precision of a measurement are different concepts as they are determined by different types of experimental errors. Thus, accuracy is a measure of how much the measured value deviates from an expected value, such that a low accuracy is likely determined by a systematic error. In turn, precision is related to the agreement of repeated experiments affected by the uncertainty of the experimental setup indicated by the data spread about the expected
value. Therefore, a repeated measurement with low accuracy but high precision will result in data values tightly grouped about an average showing a large deviation from the expected value.

Failing to account carefully for the sources and types of error can hinder the correct interpretation of experimental results. For instance, data features produced by systematic errors may be confused for physical phenomena, whilst legitimated discoveries may be conversely attributed to systematic errors. This is why the reproducibility of experimental results by different research groups is a highly priced validation criterion and practice in science.

B. Calculations

As we have noticed above, a respectable experimental scientific report ought to specify errors. This can be done either numerically or using error bars on graphical renderings of data. There are quite a few formulas that can be employed to capture the uncertainty of experimental measurements.

Percent Error: An experiment can be used to estimate numerically a quantity and then compare it to a putatively precise expected value. The expected value may be a generally accepted number (like a universal constant), or may be calculated using a mathematical expression, or may be the average (or mean) of \( N \) repeated measurements \( x_{1,2,3...N} \) of the quantity:

\[
\bar{x} = \frac{1}{N} (x_1 + x_2 + x_3 + ... + x_N)
\]

(1)

In this case, denoting \( X \) the expected value and \( x \) the experimental result, the adequacy of the measurement can be quantified using the percent error:

\[
PE = \left( \frac{x - X}{X} \right) \times 100\%
\]

(2)

Mean Deviation: Finding the actual sources of random errors polluting the data of an experiment may be a rather hopeless task, but their impact can be fairly easy to quantify owing to their random nature. Thus, when \( N \) measurements are repeated in identical conditions, the data \( x_{1,2,3...N} \) tend to disperse about the mean value \( \bar{x} \). Then, the uncertainty can be characterized by calculating the mean deviation:

\[
MD = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| = \frac{1}{N} (|x_1 - \bar{x}| + |x_2 - \bar{x}| + ... + |x_N - \bar{x}|).
\]

(3)

Standard Deviation: An even better measure of experimental uncertainty is the so called standard deviation which – when based on sufficient repeats – estimates the maximum spread of values about the mean:

\[
SD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_N - \bar{x})^2}{N}}.
\]

(4)

C. Reporting

Once the mean deviation or standard deviation is calculated, it is customary to report the result \( x \) of the respective measurement in one of the following forms which indicate the precision of the respective measurement using one of the deviations about the mean:

\[
x = \bar{x} \pm MD \quad \text{or} \quad x = \bar{x} \pm SD.
\]

(5)

The deviation may also be reported as a percent of the mean (relative error):

\[
x = \bar{x} \pm \frac{MD}{\bar{x}} 100\% \quad \text{or} \quad x = \bar{x} \pm \frac{SD}{\bar{x}} 100\%.
\]

(6)

**Example 2:** Suppose that, in the friction experiment described earlier, you measured the stopping distance travelled by the sliding block 5 times for each face. The data are tabulated on the right. Then, the mean value of the distance is

\[
\bar{x} = \frac{1}{5} (42 + 44 + 40 + 41 + 38) = 41 \text{ cm}
\]

Thence, using the proper significant figures, the mean deviation and standard deviation are about the same:

\[
MD = \frac{1}{5} (1 + 3 + 1 + 0 + 3) = 2 \text{ cm}
\]

\[
SD = \sqrt{\frac{1}{5} (1^2 + 3^2 + 1^2 + 3^2)} = 2 \text{ cm}
\]

Therefore, the various forms the distance can be reported either as \( 41 \pm 2 \text{ cm} \) or \( 41 \text{ cm} \pm 5\% \). To verify the hypothesis, the distances measured on the other two faces of the block should be comparable within an error of about 2 centimeters.
As discussed below, most Physics articles report experimental findings using graphs, inasmuch as plotting data can prove useful not only for reporting per se, but also as a heuristic tool. Accordingly, graphical representations will be ubiquitous in this lab and you will be kindly required to review a modicum of basic rules for data plotting:

1. Choose axis scales consistent with the scale of the represented data sets, such that the graph will be commensurate with the size of the frame. (For instance, a spatial axis shouldn’t be scaled in tens of centimeters if the represented data is in millimeters; the graph will look too small.)

2. Title the plot, label the axes meaningfully, specify units, and use a legend if the graph includes more than one curve.

3. After the data points are plotted, draw a smooth line indicating the trend of the data distribution. That is, avoid connecting point with the line: recall that the data points deviate from their mean with equal probability on each side, and the line should be indicative for the mean. So the line should be a curve of “best fit.”

4. The errors can be represented on the graph using error bars with length given by the deviation associated with each point.

**4. Graphing Experimental Data**

Graphs illustrate the mutual dependency between at least two physical quantities typically using a system of axes. Rather than multiple dependencies, in our class we shall study the variation of only one quantity with respect to another, so the graphs will have only two perpendicular axes. The particular dimension and scale of the represented quantities will dictate the formatting of each axis. Ordinarily, we shall use the chart tools in Microsoft Excel.

**D. Principle and Design**

A *rectangular system* of axes includes two perpendicular axes: a horizontal one called *abscissa* and a vertical one called *ordinate*. When representing experimental data, the independent variable *x* should be associated with the abscissa, while the dependent variable *y* will take values placed on the ordinate. Then, for a certain measured data pair (*x*,*y*), a data point can be found at the point of intersection of perpendiculars raised from the respective values: that is, *x* and *y* are the coordinates of the respective data point. The data points will form a discrete distribution where each point can be read as the response *y* of the natural mechanism under study when *x* takes a certain value. Thus, the multiple-point plot represents the *evolution* or *functionality* of the respective mechanism when *x* changes.

So, within the limits of experimental accuracy and precision, the graph mirrors an objective behavior. When representing a natural phenomenon, the denser is the distribution of points, the closer is the respective representation to the otherwise continuous change of the real system. As we discussed in the previous lab, error will also affect the data points which – depending on the amplitude of the random errors – will spread about an average within a range proportional to the uncertainty of that particular measurement. This average can be approximated with a continuous curve of “best fit” passing through the distribution such that the data points are evenly distributed on sides. This is the experimental result to be contrasted with the predictions made by the tested hypothesis about the respective phenomenon.

When reporting the results, special attention must be dedicated to formatting the graph for readability and scaling and labeling. Here are a few pointers:

1. *Scale* the axes to frame the range and precision of your data appropriately. For instance, if the variable is a length that takes values between 1.5 cm and 7.3 cm, the axis should be scaled in centimeters with minimum and maximum bounds at 0 and 10 cm, labeled major tick-marks each centimeter, and millimeter minor tick-marks.

2. *Label* each axis according to the associated quantity and specify units in brackets.

3. *Title* the graph meaningfully. Especially in formal reports, graphs are expected to be self-sufficient in explaining their content, so titles are often accompanied (or even replaced) by labels (such as in the Examples below) followed by captions conveying succinct descriptions of the graphed behavior or some salient feature on the graph.

4. Include a *legend* if more than one plot is represented on the same graph.
**Example 3:** Suppose that, in the spirit of our future lecture discussion about one-dimensional kinematics, you decide to represent graphically the motion of a toy car. Experimentally, this can be addressed by collecting pairs of position-time values, maybe pulling the toy along a meter stick and recording positions and times every second. You may repeat the measurement a few times to quantify the uncertainty of your experimental setup using mean deviation. The 10-second data may look like in the adjacent table.

Then, following the rules outlined above, you could plot the data, like in the shown graph. Note that the distribution is well framed and the scale is consistent with the uncertainty of about 1-2 cm.

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Position (cm)</th>
<th>MD (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>56</td>
<td>2</td>
</tr>
</tbody>
</table>

**E. Analysis**

With a modicum of ingenuity, graph can reveal many aspects of the represented phenomenon. For instance, the curve of best fit of the data-point distribution may be compared to the mathematical model suggested by an *a priori* scientific hypothesis for confirmation, or it can be conceptualized into an *empirical* model based solely on the result of the experiment. Moreover, the best fit can be used to find various characteristics of the system or phenomenon under scrutiny.

In our class we won’t actually calculate best fit curves (done via a mathematical procedure called *regression*). Instead we are to approximate the curves by drawing smooth lines through data-point distributions representing their trend inside the range of uncertainty represented using error bars, as explained in section 3C above. Therefore, it will be to our advantage to deal with linear distributions which are easier to analyze. A linear fit satisfies the equation

\[ y = b + mx, \]  

where \( m \) is the slope of the respective line and \( b \) is called the intercept because it correspond to the \( y \)-value where the graph intersects the ordinate axis (that is, where \( x = 0 \)). So, for a given linear distribution, one can calculate the experimental slope by selecting two arbitrary points on the best fit with coordinates \((x_1, y_1)\) and \((x_2, y_2)\). Then

\[ m = \frac{y_2 - y_1}{x_2 - x_1}. \]  

**Example 4:** When corroborated with the kinematic model of one-dimensional motion, the graph built above can be readily employed to find the velocity of the toy car. In class we learned that the velocity is given by the slope of the curve in position vs. time representation:

\[ v = \frac{\Delta x}{\Delta t}. \]

So, we may proceed as described above:

1. Draw a smooth line through the distribution.
2. Select two points on the line located conveniently, so the coordinates can be easily read: say for instance \((x_1, y_1) = (0,0)\) and \((x_2, y_2) = (5,30)\).
3. Use the coordinates to calculate the slope:

\[ v = \frac{30 \text{ cm} - 0}{5 \text{ s} - 0} = 6.0 \text{ cm/s}. \]
Fortunately, many mathematical models can be reduced to linear dependencies. Here is an example that will come in handy in our coming experiment: Say that the dependent variable $y$ is given theoretically by a constant $m$ times an arbitrary power of the independent variable $x$:

$$y = b + mx^n$$

The graph corresponding to this function is linear only for $n = 1$. However, we can still take advantage from the simplicity of a linear plot by redefining the variable $X = x^n$ and plot $y = b + mX$.

4. **Pre-lab**

A. **Activities**

1. Peruse carefully through the introductory material provided above. In case a certain concept is not clear enough, feel free to contact your lab-instructor with any question that you may have. Make sure that you understand
   - the two types of errors that can occur in measuring and how to compensate or correct for each type of error
   - the differences between accuracy and precision
   - how to compute the three types of errors presented in Subsection 3B. You will be using these tools to analyze your data which will help when formulating your conclusions
   - the elements required to format a stand-alone graph
   - calculating the slope of a linear graph and the logic behind turning a nonlinear function into a linear one

2. Read the preliminary information provided below and riffle through the associated Lab Form to familiarize yourself with the theoretical background of the experiment and the lab tasks that you and your team are to eke through during the next lab session.

4. Answer the questions on the quiz at the end of this document. The questions are also available on the Blackboard site associated with the PHYS-154 lecture (not lab), and you are to use that system. Only in special situations such as when you don’t have access to internet when needed, you may answer the questions on paper and turn them in at the beginning of the lab class.

5. Bring your laptop to lab next week.

B. **Preliminary information**

The subject of our first lab is “Experimental Errors and Data Analysis”. It will give you an opportunity to practice some of the concepts introduced above. To that end, we shall imagine and analyze a concrete physical situation presented as a scenario (see the Lab Form) involving people swinging out on a rope. You will be required to come with a hypothesis and then test it regarding the dependency of the swinging time on the weight of the dangling person and rope length.

To analyze the scenario, you will set up an experiment using a simple pendulum meant to model the people swinging out at the end of a rope. Note that the subject of the experiment is not the pendulum per se, but the method of data analysis, so you don’t have to worry about the rationale of the theory behind the behavior of this device.

**Pendulum Concepts**:

- The time it takes for the pendulum to complete a swing back and forth is called the *Period* ($T$).
- The distance from the pivot point to the *center of mass* of the pendulum bob is the *Arm length* ($L$)
- The *displacement* relative to the vertical can be obtained either by measuring the deflection angle ($\theta$) or the height ($h$)
above the lowest point the bob passes through. These can be related through right triangle trigonometry via

\[ h = L(1 - \cos \theta). \]  

(10)

- A theoretical model for the time period \( T \) of a simple pendulum can be approximated using the following formula valid when the displacement angle is small (less than or equal to 20°):

\[ T = 2\pi \sqrt{\frac{L}{g}}, \]  

(11)

where \( g \) is the gravitational acceleration, \( g = 9.8 \text{ m/s}^2 \)

- Note that, according to this formula, the mass of the swinging people shouldn’t influence the period as long as the displacement angle is small – this is relevant for PART 1 of the experiment.

- Moreover, in PART 2 of the experiment you will have to validate this model. Because according to this formula the dependency of the period on arm length is not linear, you will have to linearize the expression as following:

\[ T^2 = \left( \frac{4\pi^2}{g} \right) L. \]  

(12)

- This expression suggests that the slope of the plot of \( y = T^2 \) vs \( x = L \) is related to the gravitational acceleration \( g \), so you can check out the validity of your data by comparing the experimental value of \( g \) to the expected value of 9.8 m/s².

- Note that this is not the only way to redefine variables in Equation (11) to obtain a linear relationship. Think about this when answering question Q5 on the pre-lab quiz.

(Quiz on the next page)
C. Quiz 1

Based on the pre-lab readings, please answer the following questions on the Blackboard site associated with our PHYS 154 lecture. The numbers in square brackets indicate the points allotted for the respective question.

Q1. [2] In scientific research, a theory is expected to be tested experimentally by more than one research group because
   a) this may eliminate the impact of hidden systematic errors.
   b) this may eliminate the impact of random errors.
   c) to be acceptable, experimental results must be reproducible.
   d) scientific theories lose validity in time, so they must be reconfirmed periodically.
   e) Both (a) and (c) are true.

Q2. [2] Consider the following data set representing a quantity measured repeatedly 4 times: {3.2, 3.1, 3.3, 3.2}. Say that, based on a theoretical model, one calculated the expected value for the respective quantity to be 3.2. What are the percent error of the set average relative to the expected value, and the mean deviation of the set?
   a) 0.05 and 0
   b) 0 and 3.2
   c) 3.2 and 0.05
   d) 0 and 0.05
   e) None of the above.

Q3. [2] In the equation \( y = b + mx \) the quantities \( y, x, b \) and \( m \) are, in this order,
   a) independent variable, dependent variable, \( y \)-intercept, and slope.
   b) dependent variable, independent variable, \( y \)-intercept, and slope.
   c) \( y \)-intercept, independent variable, \( x \)-intercept, and slope.
   d) \( y \)-intercept, independent variable, slope, and mass.
   e) dependent variable, independent variable, slope, and mass.

Q4. [2] In our lecture we learned that the position \( x \) of a particle moving in a straight line with constant velocity \( v \) is given by the linear equation \( x = x_0 + vt \). Say that such a motion is given by the adjacent chart. Based on the graph, what are the initial position \( x_0 \) and velocity \( v \) of this particle?
   a) 4.0 m, -2.0 m/s
   b) 2.0 m, 3.0 m/s
   c) 2.0 m, 2.0 m/s
   d) 0.0 m, 2.0 m/s
   e) None of the above.

Q5. [2] Which of the following changes of variable will linearize the equation of a pendulum period \( T \) in terms of its length \( L \)?
   a) \( y = T^2, x = L \)
   b) \( y = T, x = \sqrt{L} \)
   c) \( y = T^2, x = L/g \)
   d) All of the above.
   e) None of the above.