ERROR ANALYSIS EXPERIMENT

Introduction

In this experiment we are interested in learning how to treat data. The main point here is that any time a scientist measures a number; he/she must be able to tell his colleagues the accuracy of that number. Otherwise, there is no way to tell whether the number agrees with the predictions of a theory; there is also no way for another scientist to check the experiment. For example, suppose I measure the circumference of a circle, then its diameter, and divide the circumference by the diameter. The result ought to be $\pi$. If my result is 3.15, have I proved that $\pi$ is not 3.14?

Equipment

Vertical rod, clamp, hook-rod, string (at least 1m), mass hanger, mass (100 g), metal block.

Procedure

PART 1 - Reading Scale

The most common measurements in a lab are done with devices that have a marked scale. Use the ruler to measure the length of the three sides of the metal block on the table.

<table>
<thead>
<tr>
<th>X-side (mm)</th>
<th>Y-side (mm)</th>
<th>Z-side (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The length obtained from a single measurement can be written as $L = L_0 \pm \delta L$ where $L_0$ is the value measured for any of the sides ($X, Y$ or $Z$) and $\delta L$ is the error of each single of these measurements. $\delta L$ is defined by the accuracy of the instrument used. What is $\delta L$ of the ruler (the smallest marked interval)?

1) $\delta L = \boxed{\phantom{0}}$

Chose one face of the metal block and calculate the perimeter $P_0$ of it.

$P_0 = \boxed{\phantom{0}}$

What about the error of the measurement of the perimeter? Make a prediction, would the error be the same/greater/smaller than $\delta L$?

The correct way to calculate the error of the perimeter $\delta P$ is given by (error propagation for the sum of independent variables):

$$\delta P = \sqrt{4 \cdot (\delta X)^2 + 4 \cdot (\delta Y)^2}$$

Use the previous equation to calculate $\delta P$, where in our case $\delta X = \delta Y = \delta L$

$\delta P = \boxed{\phantom{0}}$
So the final answer for the perimeter is

2) \( P = P_0 \pm \delta P = _____ \pm _____ \)

Next calculate the volume of the metal block:

\( V_0 = _____ \)

What about the error of the volume \( \delta V \)? Make a prediction, would the error be the same/greater/smaller than three times \( \delta L \)?

The correct way to calculate the error of the volume \( \delta V \) is given by (error propagation for the product of independent variables):

\[
\delta V = V_0 \sqrt{\left(\frac{\delta X}{X}\right)^2 + \left(\frac{\delta Y}{Y}\right)^2 + \left(\frac{\delta Z}{Z}\right)^2}
\]

Use the previous equation to calculate \( \delta V \). In our case \( \delta X = \delta Y = \delta Z = \delta L \)

\( \delta V = _____ \)

So the final answer for \( V \) is:

3) \( V = V_0 \pm \delta V = _____ \pm _____ \)

**PART 2 – Average and Standard deviation**

In this lab you will measure the period of oscillation, \( T \), of a pendulum. Use the computer to find a webpage with a stopwatch. The online stopwatch has its own precision (use at least \( \pm 0.01 \) s), but there is a difference with PART 1. We now have a more significant source of error.

4) What would it be?

Our best estimate of the ‘real’ value is the average. Measure \( T_i \) twenty times (with accuracy of \( \pm 0.01 \) s, round it if needed. Note: the stopwatch on your phone might have an accuracy of \( \pm 0.1 \) s which is not precise enough) and use the next table to collect your data. You will learn later in the semester that \( T \) depends only on the length of the pendulum, not on the value of the hanging mass or the amplitude (how far away from the equilibrium position you displace the pendulum). But you need to keep this displacement less than about 15 degrees.

5) \( \bar{T} = \frac{1}{N} \sum_{i=1}^{N} T_i = _____ \) (Use at least four digits for the average)

where \( N \) is the number of the measurements taken. Next, what is the error of our measurement? First let’s introduce \( d \), the deviation from the average value:

\( d_i = T_i - \bar{T} \)

Calculate \( d_i \) for each of your measurements.
We are interested in the average deviation from our best estimate. So should we just take the average of the $d_i$'s?

6) $\bar{d} = \frac{1}{N} \sum_{i=1}^{N} d_i = \ldots$

Clearly, the average of deviations cannot be used as the error estimate, since it gives us zero (or very close to zero, depending on how you have rounded your digits). This is because the deviations with positive sign are always canceled by the deviations with negative sign. In order to evaluate the error of the average $\sigma_{\text{average}}$ first we need the standard deviation $\sigma_T$. It’s a "natural" measure of the statistical dispersion when the center of the data is measured about the average.

Calculate the standard deviation:

7) $\sigma_T = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} d_i^2} = \ldots$

The error for the average $\sigma_{\text{average}}$ is related to the standard deviation $\sigma_T$ by:

$\sigma_{\text{average}} = \frac{\sigma_T}{\sqrt{N}}$

So the final answer for $T$ is:

8) $T = \bar{T} \pm \sigma_{\text{average}} = \ldots \pm \ldots$

Looking at the definition of $\sigma_{\text{average}}$, what can be done to reduce it?
PART 3 – Data Distribution

We now want to display the data in the form of a histogram to show the frequency distribution of your measured times. This will be somewhat like a histogram that shows the distribution of student grades on an exam where, for example, the histogram shows the number of grades in 10 point intervals (bins) from 0 to 100.

We will use Excel to plot the histogram. Enter your 20 measurements of \( T_i \) in column A in Excel. In column B enter bin numbers. These numbers will span the range of your measurements from low to high in increments of 0.05 s. For example, if your lowest \( T_i \) was 1.59 s and your highest \( T_i \) was 1.82 s, then your bin numbers would be 1.60 s, 1.65 s, ..., 1.80 s, 1.85 s. Excel will determine the number of values of \( T_i \) in each bin where the bin number represents the largest value of \( T_i \) in each bin.

Click the Data tab and then Data Analysis and select Histogram from the menu. The Input Range is the range of cells representing the data (column A) and the Bin Range is column B.

Click Chart Output and then OK and Excel will plot the histogram of your data. Select and print your histogram plot.

Sketch a bell shaped curve on your graph that would correspond to a rough visual fit to the data. Ideally, if you took enough measurements this distribution would be given by a normal or Gaussian distribution function, like that shown below.

\[
\begin{align*}
\text{n} & \quad \text{T} \\
-\sigma_T & \quad \bar{T} & \quad +\sigma_T
\end{align*}
\]

This curve helps to visualize the average value and the spread about the average.

9) How is the peak position of your curve related to the average value you calculated in part 5?

The Gaussian distribution also provides a graphical meaning for the standard deviation. Mark with a dot on the x-axis the average value of \( T \). Mark also on the x-axis the two values of the standard deviation: \(-\sigma_T\) one on the left and \(+\sigma_T\) on the right of the average of \( T \). Draw two vertical lines (dotted lines) intersecting the x-axis at \( \pm\sigma_T \). Your graph should look like the figure above.

10) How many of your data points lies between \( \pm\sigma_T \)?

11) Which percentage does it correspond to (the total number of data is \( N=20 \))?

Ideally, you should have obtained 68%. This 68% of values within \( \pm\sigma_T \) is a mathematical property of the Gaussian distribution.

12) If you take the standard deviation as a measure of how well you can determine \( T_i \) in a single measurement, how does it agree with the initial guess of 0.01 s?