Experimental error and data analysis

Purpose
To review the types of experimental errors and some methods of error and data analysis that will be used in subsequent experiments.

Theory
Any measurement of physical quantities always involves some uncertainty or experimental error. One should not only report a result but also give some indication of its uncertainty of the experimental data.

1. Types of experimental errors
(a) Random or statistical errors result from unknown and unpredictable variation that arise in all experiment situations, for example, fluctuation in temperature or voltage, mechanical vibrations of an experimental setup, unbiased estimates of measurement readings by the observer. Repeated measurements with random error give slightly different values each time. The random error can be estimated by repeating an experiment several times.
(b) Systematic errors are associated with particular measurement instruments or techniques, for example, an improperly calibrated instrument or personal error, such as using a wrong constant in a calculation or always taking a low reading of a scale division. Avoiding systematic errors depends on the skill of the observer to recognize the sources of such error and to prevent or correct them.

2. Accuracy and precision
Accuracy and precision are commonly used synonymously, but in experimental measurements there is an important distinction.
(a) Accuracy signifies how close the measured value comes to the true or accepted value, that is, how correct it is.
(b) Precision refers to the agreement among repeated measurements or how close they are together.

3. Significant figures
The significant figures of an experimentally measured value include all the numbers that can be read directly from the instrument scale plus one doubtful or estimated number (fraction of the smallest division). For example, the length of an object may be read as 2.54 cm (three significant numbers) on one instrument scale and 2.5405 cm (five significant numbers) on another. Thus, the significant numbers depend on the quality of the instrument and the fineness of its measuring scale.
4. **Computations with measured values**
   
   (a) **Addition and subtraction:** Begin with the first column from the left that contains a doubtful figure, round off all numbers to this column and drop all digits to the right.

   \[
   42.31 \ 0.0621 + 512.4 + 2.57 = 42.3 + 0.1 + 512.4 + 2.6 = 557.4
   \]

   (b) **Multiplication and division:** The number of significant figures in the final answer in general equals the number of significant figures in the measurement with the least number of significant figures.

   \[
   6.27 \times 5.5 = 34.485 \quad \text{(from a hand calculator)} = 34.37429
   \]

   \[
   \frac{374}{29} = 12.896551 \quad \text{(from a hand calculator)} = 13.5.
   \]

5. **Expressing experimental error and uncertainty**

   (a) **Percentage error**

   The accepted or "true" value of a physical quantity found in textbooks and handbooks is the most accurate value through sophisticated experiments or mathematical methods.

   The **absolute difference** between an experimental value \( x \) (defined in 5(b)) and the accepted value \( A \), is written in \( |x - A| \). The percentage error of the experimental value is:

   \[
   \text{Percentage error} = \frac{\text{absolute difference}}{\text{accepted value}} \times 100\% = \frac{|x - A|}{A} \times 100\% 
   \]

   The accuracy of the experimental value is expressed in terms of percentage error.

   (b) **Average (Mean) value ( \( \bar{x} \) )**

   The average or mean value of a set of \( N \) measurements is:

   \[
   \bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i
   \]

   (c) **Deviation from the mean**

   Having obtained a set of measurements and determined the mean value, it is helpful to report how widely the individual measurements are scattered from the mean. A quantitative description of this scatter or dispersion of measurements will give an idea of the precision of the experiment.

   - **Deviation** \( d_i \)
     
     The deviation \( d_i \) from the mean of any measurement with a mean value \( \bar{x} \) is

     \[
     d_i = x_i - \bar{x}.
     \]

   - **Mean deviation** \( \bar{d} \)
     
     The mean deviation \( \bar{d} \) is a measure of the dispersion of experimental measurements about the mean, i.e., a measure of precision.

     To find the mean deviation \( \bar{d} \) of a set of \( N \) measurements, the absolute deviations \( |d_i| \) are determined first,

     \[
     |d_i| = |x_i - \bar{x}|
     \]
The **mean deviation** $\bar{d}$ is then
\[ \bar{d} = \frac{|d_1| + |d_2| + |d_3| + \ldots + |d_N|}{N} = \frac{1}{N} \sum_{i=1}^{N} |d_i| \]

**Example:** What is the average $\bar{x}$ or mean value of the set of numbers 5.42, 6.18, 5.70, 6.01, and 6.32? What is its mean deviation $\bar{d}$?

\[ \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{5.42 + 6.18 + 5.70 + 6.01 + 6.32}{5} = 5.93 \]
\[ |d_1| = |5.42 - 5.93| = 0.51 \]
\[ |d_2| = |6.18 - 5.93| = 0.25 \]
\[ |d_3| = |5.70 - 5.93| = 0.23 \]
\[ |d_4| = |6.01 - 5.93| = 0.08 \]
\[ |d_5| = |6.32 - 5.93| = 0.39 \]
\[ \bar{d} = \frac{1}{N} \sum_{i=1}^{N} |d_i| = \frac{0.51 + 0.25 + 0.23 + 0.08 + 0.39}{5} = 0.29 \]

It is common to report the experimental value of a quantity in the form: $\bar{x} \pm \bar{d}$. In the above example, this would be 5.93 ± 0.29 or 5.9 ± 0.3.

**An experimental value and its mean deviation should have their last significant digits in the same location relative to the decimal point.** The ± gives a measure of the precision of the experimental value.

It is also common practice to express the dispersion on the mean deviation as a percent of the mean:
\[ \bar{x} \pm \frac{\bar{d}}{\bar{x}} \times 100\% \]

For the above example, we have 5.93 ± \( \frac{0.29}{5.93} \times 100\% = 5.93 \pm 4.9\% \)

**(d) Standard deviation** (optional for PHY 2048L)

Standard deviation, $\delta$, is defined as:
\[ \delta = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \]

**(e) Propagation of errors** (optional for PHY 2048L)

- **Addition and subtraction of experimental values**
  Suppose $x$, $y$, and $z$ are three measured values and the errors are $\delta_x$, $\delta_y$, $\delta_z$. If $w$ is the value to be calculated from these measurements and is defined to be:
  \[ w = x - y + z \]
  Then statistical analysis shows that in good approximation the error $\delta_w$ is:
  \[ \delta_w = \sqrt{\delta_x^2 + \delta_y^2 + \delta_z^2} \]

- **Multiplication and division of experimental values**
  Suppose $x$ and $y$ are two measured values and the errors are $\delta_x$ and $\delta_y$. If $w$ is the value to be calculated from these measurements and is defined to be:
\[ w = x \times y \text{ (or } w = \frac{x}{y}) \]

Then statistical analysis shows that the fractional error of \( w \) is:

\[
\frac{\delta w}{w} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2}
\]

6. **Graphical representation of data**

It is often to represent experimental data in graphical form, not only for reporting, but also to obtain information.

Quantities are commonly plotted using rectangular Cartesian axes (x and y). The location of a point on the graph is defined by its coordinates \( x \) and \( y \), referenced to the origin \( O \).

When plotting data, choose axis scale so that most of the graph paper is used and the scale units should always be included.

When the data points are plotted, draw a smooth line with an approximately equal number of points on each side to make a "curve of best fit".

In case where several determinations of each experimental quantity are made, the average value is plotted and the mean deviation may be plotted as error bars.

Figure 1 shows an example of a properly labeled and plotted graph with error bars indicating mean deviation.

![Figure 1](image-url)