Appendix A

Experimental Data Analysis

The purpose of physical experiments is to discover or verify relationships among the experimental variables. In order to accomplish this goal, it is necessary to be able to estimate the uncertainty or error involved in the experimental measurements. After an error analysis, the problem then becomes one of finding or displaying most clearly the relationship between variables (taken two at a time). This process is usually best accomplished by graphical analysis of the experimental data.

A.0.1 Error Analysis

There are two general types of errors: systematic errors and random errors. Systematic errors may occur due to faulty methods or faulty apparatus. These systematic errors are usually constant during an experiment and they can be assigned a sign and magnitude once discovered. For example they may be a “zero error” on a vernier caliper or electrical meter. These imply definite zero errors in all subsequent measurements. Careful analysis of experimental procedures and apparatus is the best safeguard against systematic errors.

Random errors are the uncontrollable errors that arise in an experiment because of varying experimental conditions involving observer, apparatus, and systems under study. These random errors account for the distribution of experimental results when repeated measurements are made. These random errors are usually assumed to have normal or Gaussian distributions. These errors have equal probability of being positive or negative and are more likely to be small than large. The effect of these errors can be minimized by taking a large number of measurements and using average values as best estimates of true values.

Measurements with Equal Uncertainties.

If a number of measurements of a quantity \( x \) have been made, and if each individual measurement has the same uncertainty as any other, then;
\[ \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \]  
(A.1)

There are various ways of indicating error limits for a measured quantity. The number of significant figures is referred to as a first approximation error analysis. If \( x = 5 \), it is meant that \( x \) could be any value between 4.5 and 5.5. If \( x = 5.0 \) then \( x \) lies between 4.95 and 5.05. Power of ten notation allows precision in significant number notation, e.g., \( 5 \times 10^3 \) and \( 5.00 \times 10^3 \) clearly indicate the precision needed.

A second order approximation in error analysis is to use the difference between maximum and minimum measurements of a quantity as error range of measurements. For example, length measurements ranging between 5.0 and 5.4 cm with an average of 5.2 cm would be expressed as \( x = 5.2 \pm 0.2 \text{ cm} \).

Another second order approximation is the use of mean deviation. The mean deviation is defined as the average deviation of the measurements from the mean value of the measurements;

\[
\text{mean deviation} = \frac{\sum_{i=1}^{N} |x - x_i|}{N} \quad \text{(A.2)}
\]

For example, given the five time measurements; \( t_1 = 5.0 \text{ sec.}, \ t_2 = 5.2 \text{ sec.}, \ t_3 = 4.9 \text{ sec.}, \ t_4 = 5.4 \text{ sec.}, \) and \( t_5 = 5.0 \text{ sec.} \),

\[
\bar{t} = \frac{5.0 + 5.2 + 4.9 + 5.4 + 5.0}{5} = 5.1 \text{ sec.}
\]

mean dev. = \( \frac{1 + 1 + 0.2 + 0.3 + 1}{5} = 0.2 \text{ sec.} \)

A third approximation in error analysis involves using the distribution of data. Under the conditions of random errors and normal or Gaussian distribution (the common bell shaped distribution curve shown in figure A.0.1) - powerful mathematical analysis of errors is possible.

The standard deviation for such a Gaussian distribution is defined as follows:

\[
\sigma = \left[ \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N} \right]^{\frac{1}{2}} \text{ for large } N \text{ values} \quad \text{(A.3)}
\]

and the probability of finding \( x \) values in the following ranges are calculated:

\[
\bar{x} - \sigma \leq x \leq \bar{x} + \sigma \quad \text{probability} = 0.683
\]

\[
\bar{x} - 2\sigma \leq x \leq \bar{x} + 2\sigma \quad \text{probability} = 0.955
\]

\[
\bar{x} - 3\sigma \leq x \leq \bar{x} + 3\sigma \quad \text{probability} = 0.997
\]

Since 2/3 of the data points fall within ±\( \sigma \) of \( \bar{x} \), \( \sigma \) can be used as an error estimate and then \( x = \bar{x} \pm \sigma \). Example: Using the previous data and the small \( N \) approximation we find
Figure A.1: Gaussian or Normal Distribution

\[ \sigma = \left[ \frac{(0.1)^2 + (0.1)^2 + (0.2)^2 + (0.3)^2 + (0.1)^2}{4} \right] = 0.2 \text{ sec.} \]

and

\[ \bar{x} = 5.1 \pm 0.2 \text{ sec.} \]

**Measurements with Unequal Uncertainties**

If a number \((N)\) of measurements of a quantity \((x)\) have been made, but there are different uncertainties \((\sigma_i)\) associated with each measurement \((x_i)\), then it is proper to calculate a weighted mean for \((x)\):

\[ \bar{x} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \quad (A.4) \]

and an estimated error of the mean:

\[ \sigma_x = \frac{1}{\sqrt{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}} \quad (A.5) \]

**A.0.2 Error Propagation**

In many experiments, the determination of a quantity is a result of a computation using a known relationship between measured quantities and the dependent quantity, e.g., density = mass/volume or \(\rho = m/V\), and the density is computed from mass
and volume measurements. The problem of error analysis becomes one of error propagation due to error in measurements of \( m \) and \( V \). The procedure for deriving such an error equation is as follows:

- Take the natural logarithm of both sides of the equation: \( \ln \rho = \ln m \) – \( \ln V \)
- Take the differential of this equation: \( \frac{\Delta \rho}{\rho} = \frac{\Delta m}{m} - \frac{\Delta V}{V} \)

(This is a determinate error equation – in order to convert it an indeterminate error equation all signs should be chosen to make the error estimate a maximum.)

Therefore:

\[
\frac{\Delta \rho}{\rho} = \frac{\Delta m}{m} + \frac{\Delta V}{V} \tag{A.6}
\]

where \( \Delta m = \text{error in } m; \Delta V = \text{error in } V; \bar{m} = \text{average value of } m; \bar{V} = \text{average value of } V; \) and \( \frac{\Delta \rho}{\rho} = \text{fractional error in density.} \)

Usually, however, we do not know the actual errors in the parameters, but we have a characteristic uncertainty or estimated error in each parameter, expressed as a standard deviation (\( m \) and \( \sigma_m \)) as well as the volume and its standard deviation (\( V \) and \( \sigma_V \)). To combine:

\[
\sigma_{\rho}^2 = \sigma_m^2 \left( \frac{\partial \rho}{\partial m} \right)^2 + \sigma_V^2 \left( \frac{\partial \rho}{\partial V} \right)^2 \tag{A.7}
\]

which is valid as long as the uncertainties in \( m \) and \( V \) are uncorrelated. This should be true if the measurements which were performed to determine the mass and the volume were completely independent of each other. In our case this reduces to simply:

\[
\left[ \frac{\sigma_{\rho}}{\rho} \right]^2 = \left[ \frac{\sigma_m}{m} \right]^2 + \left[ \frac{\sigma_V}{V} \right]^2 \tag{A.8}
\]

Equations (A.7) or (A.8) will be used in almost all of the error propagation that you will have to perform in the laboratory. For the most common combinations of variables by multiplication or division (as in the above example) or addition or subtraction, it reduces to the familiar equations:

If \( z = x + y \) or if \( z = x - y \) then:

\[
\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{A.9}
\]

If \( z = xy \) or if \( z = \frac{x}{y} \) then:

\[
\frac{\sigma_z}{z} = \sqrt{\left( \frac{\sigma_x}{x} \right)^2 + \left( \frac{\sigma_y}{y} \right)^2} \tag{A.10}
\]
A.0.3 Graphical Analysis

After error analysis is made, it may be that the problem is one of finding the relationship between the measured variables. The most efficient way to do this is to make a graphical analysis of the data. The following guidelines are provided to assist your graphical analysis.

1) The dependent variable is always plotted along the vertical \( y \) axis and the independent variable along the horizontal \( x \) axis: e.g., the heart rate is a function of physical activity and heart rate would be plotted on the \( y \) axis and physical activity would be plotted along the \( x \) axis.

2) Error bars may be used to indicate errors in measurements.

3) Linear relationships can be very accurately determined by graphical analysis. It is therefore desirable to find the relationship that yields the linear relation that is indicated by a straight line plot, e.g., consider the study of falling body velocity as a function of initial height.

![Figure A.2: Linear & Non-Linear Relations](image)

Semilog graph paper produces straight line plots for exponential relations: \( y = ar^x; \)
\( a = \text{constant}; r = \text{constant}. \)

\[
\log_{10} y = \log_{10} a + x \log_{10} r
\]
i.e.,
\[
\log_{10} y = mx + \log_{10} a
\]
where
\[
m = \text{slope} = \log_{10} r
\]
APPENDIX A. EXPERIMENTAL DATA ANALYSIS

Semilog paper has a logarithmic $y$–axis (it automatically takes logarithms of data plotted) and a regular spaced $x$ axis. The best way to find slopes from a semilog plot is to use vertical axis changes of $\Delta(\log_{10} y) = 1$ (see graph in figure A.0.3 and find the corresponding $\Delta x$; the slope is then $1/\Delta x$.

Log–Log graph paper produces a straight–line plot for power law relations. Consider $y = ax^m$; then $\log_{10} y = \log_{10} a + m \log_{10} x$. Log–Log graph paper has logarithmic scales on both $x$ and $y$ axis and automatically takes logarithms of data plotted. The slope of a log–log straight line is the actual geometrical slope (as measured with a ruler!). See figure A.0.3

**Example:** Area vs. side of cube

\[
slope = \frac{2 \text{ cm}}{1 \text{ cm}} = 2
\]

Therefore: $A = x^2$. 

Figure A.3: Semi-Log Plots

Figure A.4: Log Log Plots
A.0.4 Method of Least Squares

If $N$ measurements are made of quantity $x$ and a normal distribution is assumed for errors, then the most probable value of $x$ is the mean:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$  \hspace{1cm} (A.11)

with a standard deviation of the mean, $\sigma_x$, given in terms of standard deviations of the individual observations, $\sigma$, by the equation:

$$\sigma_x = \frac{\sigma}{\sqrt{N}}$$  \hspace{1cm} (A.12)

For the linear relation $y = mx + b$, we can find the most probable values of $m$ and $b$ from a set of measurements of $(x_i, y_i)$. If we assume negligible errors in $x_i$ and a normal distribution of $y_i$ errors, then the most probable values of $m$ and $b$ are given as follows:

$$m = \frac{N \sum_{i=1}^{N} x_i y_i - \left[ \sum_{i=1}^{N} x_i \sum_{i=1}^{N} i = 1^N y_i \right]}{N \sum_{i=1}^{N} (x_i)^2 - \left[ \sum_{i=1}^{N} x_i \right]^2}$$  \hspace{1cm} (A.13)

$$b = \frac{\left[ \sum_{i=1}^{N} y_i \right] \left[ \sum_{i=1}^{N} (x_i)^2 \right] - \left[ \sum_{i=1}^{N} x_i y_i \right] \left[ \sum_{i=1}^{N} x_i \right]}{N \left[ \sum_{i=1}^{N} (x_i)^2 \right] - \left[ \sum_{i=1}^{N} x_i \right]^2}$$  \hspace{1cm} (A.14)

with

$$\sigma_m = \frac{\sqrt{N} \sigma}{\sqrt{N \sum_{i=1}^{N} (x_i)^2 - \left[ \sum_{i=1}^{N} x_i \right]^2}}$$  \hspace{1cm} (A.15)

and

$$\sigma_b = \frac{\left[ \sum_{i=1}^{N} (x_i)^2 \right]^\frac{1}{2} \sigma}{\sqrt{N \sum_{i=1}^{N} (x_i)^2 - \left[ \sum_{i=1}^{N} x_i \right]^2}}$$  \hspace{1cm} (A.16)

where

$$\sigma = \frac{1}{\sqrt{N}} \left[ \sum_{i=1}^{N} (mx_i + b - y_i)^2 \right]$$  \hspace{1cm} (A.17)
A.0.5 Chi–Squared Test

Whenever data are analyzed in terms of a functional relationship (linear, quadratic, exponential or whatever), it is important to determine just how well the data do in fact obey that relationship. In the above example of $x$ and $y$ data analyzed in terms of a linear relationship, $y = mx + b$, we may define the statistic $\chi^2$ (Chi squared) as:

$$\chi^2 \equiv \sum_{i=1}^{N} \left[ \frac{1}{\sigma_i^2} [y_i - y(x_i)]^2 \right]$$

where $x_i$ and $y_i$ are the pairs of data points, $\sigma_i$ is the uncertainty in $y_i$ and $y(x_i)$ is the value of the function, $y = mx_i + b$ calculated using the parameters $m$ and $b$ obtained from the least–squares analysis.

The “goodness of the fit”, or the extent to which the data do indeed obey the given functional relationship, is judged by the “reduced $\chi^2$” or “$\chi^2$ per degree of freedom”, $\chi^2/\nu$, where the number of degrees of freedom $\nu$ is equal to the number of data points minus the number of parameters. In the above example of fitting $N$ data points to a linear relationship with two parameters ($m$ and $b$), $\nu = N - 2$. The fit is reasonably good as long as $\chi^2/\nu$ is close to 1.0. A $\chi^2$ much larger than this indicates a poor fit, while a $\chi^2$ much smaller than this usually indicates unreasonably large error estimates ($\sigma_i$).