DATA ANALYSIS AND ERRORS OF OBSERVATION

1. INTRODUCTION

The measurement of a physical quantity is always subject to errors. A statement of the result does not mean very much unless some indication is given to the accuracy of the measurement. Some basic knowledge of error analysis is essential for any beginner.

It is customary and convenient to divide experimental errors into two types, systematic and random, according to the difference between their characters and the methods of treatment.

(a) Systematic errors are constant errors which cause all results to be incorrect by roughly the same amount in the same direction. Usually they cannot be eliminated by averaging or treated by statistical method. Care must be taken in the interpretation of systematic error.

(b) Random errors are minor experimental or accidental errors due to errors of judgment, or mildly fluctuating conditions which cannot be controlled. There are many causes of such fluctuations. Random errors may be estimated by statistical method.

In most experiments, both random and systematic errors are present. A set of measurements with random errors only are shown in Fig. 1, another set with systematic plus random errors are shown in Fig. 2. Each mark in the figures indicates the result of a measurement.
2. **SYSTEMATIC ERRORS**

Examples of systematic errors are those caused by

(a) **Incorrect calibration of an instrument**: e.g. bias caused by friction or wear in its moving parts, and the failure to correct for the zero-reading.

(b) **Construction faults in the apparatus**: e.g. off-centre pivoting of the pointer of a circular scale.

(c) **Non-constancy of experimental conditions**: e.g. change of temperature and pressure which affects the value of the quantity measured and has not been properly taken into account.

(d) **Bias of the observer**: e.g. tendency of some persons to press a stop watch a fraction of a second too early or to overestimate or underestimate the fraction of a scale division.

(e) **Incorrect use of the formulae**: e.g. the experimental arrangement is different from that assumed in the theory, and the correction factor which takes account of this difference is ignored.

(f) **Incorrect procedure or method**.

Unlike random errors, systematic errors will not be revealed or minimized just by repeating the measurements with the same apparatus. Where no independent true value is known there can be no infallible method for discovering systematic errors. There are, however, several recognized ways of tackling the problem. First we should investigate all obvious possible sources of systematic errors, such as calibrations of meters or of standards used. Second, if we suspect that some external factor may be affecting the measurement systematically, we can sometimes check this by deliberately altering the suspected factor to see if this produces any detectable change in the result.
Thirdly, we should check the validity of the formulae used in the calculation. Finally, wherever possible we should try to make the measurement by two or more methods independent of one another.

A classical example was the famous Millikan's oil-drop experiment. In the experiment Millikan measured the electronic charge \( e \) and gave the value

\[
e = (1.591 \pm 0.002) \times 10^{-19} \text{C}
\]

while the modern value is

\[
e = (1.60217733 \pm 0.00000049) \times 10^{-19} \text{C}.
\]

The discrepancy was due to some systematic error which was not recognized. It was later traced to an inaccuracy in the value of the viscosity of air used in his calculation.

3. **RANDOM ERRORS**

Even if every step possible had been taken to eliminate the systematic errors present, a quantity measured a number of times still would not give identical results. The discrepancy is attributed to random errors, so-called because it occurs without a definite pattern.

We may imagine a random error as consisting of a number of small "elementary" errors that unavoidably enter into each measurement but are not specifically recognized. Examples of elementary errors are those due to small disturbances caused by mechanical vibrations, tremors produced by the wind, misjudgement in the interpolation of the smallest division of the scale of measuring instruments, irregularities in the specimen measured, actual variation of the quantity measured due to fluctuation in the surrounding conditions, etc. The elementary errors may enhance or cancel each other and conspire to give the observed variation in the readings obtained.
On the basis of the assumption that the elementary errors are independent of each other, and that the positive and negative deviations of the same size in a variable reading are equally probable, a theoretical frequency distribution of random errors was first derived by Laplace in 1783 and is known as the Gaussian or normal law of errors.

Suppose a quantity $x$ is measured a large number of times, $n$, under supposedly identical physical conditions. Let the readings be $x_1, x_2, \ldots, x_n$. These may be represented graphically as a frequency distribution as follows. We divide the range of value in which $x$ occurs into intervals of equal size, $\Delta x$, and count the number $f$ of readings falling within each interval. A histogram is obtained by plotting $f$ against $x$ as shown in Fig. 3.

If $n \to \infty$ and $\Delta x \to 0$, the histogram becomes a smooth curve known as a distribution curve. The normal law of errors describes a distribution curve shown in Fig. 4. It is characterized by two parameters, the mean $\bar{x}$ and the standard deviation $\sigma$ which are to be defined later. The normal distribution curve is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$
Some properties of the normal distribution law may be summarized:

(a) The distribution is symmetric with respect to the mean, \( x \), which is defined by the formula

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]  

where \( n \) is the total number of measurements taken.

Physically, this means that positive and negative random errors are equally probable. A large number of independent readings of a quantity should therefore be taken. Their mean is then almost entirely free from random errors and gives the best estimate for the value of the quantity measured.

(b) For a given mean, the spread of readings is governed by the standard deviation, \( \sigma \) which is defined by the formula

\[
\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

(c) The probability \( P(T) \) that a reading \( x_i \), will occur between the readings \( \bar{x} - T\sigma \) and \( \bar{x} + T\sigma \), may be calculated theoretically.

A simple table of \( P(T) \) is given below:

<table>
<thead>
<tr>
<th>( T )</th>
<th>( P(T) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6745</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.6827</td>
</tr>
<tr>
<td>2</td>
<td>0.9543</td>
</tr>
<tr>
<td>3</td>
<td>0.9973</td>
</tr>
<tr>
<td>( \infty )</td>
<td>1</td>
</tr>
</tbody>
</table>
Geometrically, $P(T)$ corresponds to the area under the distribution curve between the boundary lines $x \pm T \sigma$. (shaded area in Fig. 5).

![Probability curve for the normal distribution in terms of $\sigma$.](image)

Fig. 5 Probability curve for the normal distribution in terms of $T = \frac{x - \bar{x}}{\sigma}$

It is seen that the probability that an observation will lie between $\bar{x} \pm 0.6745\sigma$ is 50%. In other words, the quantity $0.6745\sigma$ is the deviation from the mean which is just as likely to be exceeded as not. It is called the **probable error**. For instance, the statement that a determination of $g$ gives $981 \pm 8$ cm sec$^{-2}$, where the error indicated is the probable error, means that if 10 more identical determinations were carried out, 5 of the values obtained are expected to lie within the range between 973 cm sec$^{-2}$ and 989 cm sec$^{-2}$ and 5 values outside this range.

The probability that an observation will lie outside the range $\bar{x} - \sigma$ and $\bar{x} + \sigma$ is approximately 32%; outside the range between $\bar{x} - 2\sigma$ and $\bar{x} + 2\sigma$ is only about 5%.

(d) The mean deviation $\eta$ is estimated by the formula

$$\eta = \frac{1}{n-1} \sum_{i=1}^{n} \left| x_i - \bar{x} \right|$$

(3)
Theoretically it may be shown that $\eta$ is about 20% less than the standard deviation, $\sigma$, so that a rough but convenient estimation of the standard deviation is given by

$$\sigma \approx 1.25 \eta$$  \hspace{1cm} (4)

4. STANDARD ERROR

So far we have considered only one set of $n$ measurements of $x$, i.e. $x$ is measured $n$ times yielding results $x_1, x_2, \ldots, x_n$. Now we will extend our discussion to include many different sets of the same number of measurements of $x$ under identical conditions. The individual values of $x$ in each new set would probably be somewhat different but the mean value of each set would be much closer together than the individual values in any single set. When a number of such mean values are obtained and their deviation from the overall mean computed, a smaller dispersion or scatter of the means is expected. This overall mean is taken to represent the true value of $x$. In fact, it can be shown that, the deviation of the mean from the true value of $x$ is given by

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$  \hspace{1cm} (5)

where $\sigma$ and $\bar{x}$ are quantities belonging to the same set of $n$ measurements of $x$. $\sigma_x$ is usually called the standard error. In the literature, the error quoted for a measurement is usually the standard error. In practice, only one set of $n$ measurements is performed.

5. PROPAGATION OF ERRORS

We don't normally measure the final quantity directly, but measure several primary quantities first. For example the density of a cube, $\rho$ can be determined from its mass $M$ and its dimensions and $l_x, l_y$ and $l_z$:

$$\rho = \frac{M}{l_x l_y l_z}$$
It is apparent that the accuracy of the final value of \( \rho \) is affected by those of \( M, l_x, l_y \) and \( l_z \).

Suppose that in order to determine a physical quantity \( R(x, y, z, ...) \), we take the set of measurements \( x_1, y_1, z_1, \ldots; \ldots; x_n, y_n, z_n, \ldots \) under identical conditions. The best estimate of \( R \) is given by \( \tilde{R}(x, y, z, ...) \). How is the standard deviation of \( R \) related to the standard deviation of the directly measured quantities \( x, y, z, ... \)?

According to calculus, we write,

\[
\Delta R_i = R(x_i, y_i, z_i, \ldots ) - \tilde{R}(x, y, z, \ldots ) \\
= R(x + \Delta x_i, y + \Delta y_i, z + \Delta z_i, \ldots ) - \tilde{R}(x, y, z, \ldots ) \\
= \frac{\partial R}{\partial x} \Delta x_i + \frac{\partial R}{\partial y} \Delta y_i + \frac{\partial R}{\partial z} \Delta z_i + \ldots \tag{6}
\]

where the variations \( \Delta x_i, \Delta y_i, \Delta z_i \) are assumed to be small, and all the partial deviations are evaluated at \( x = x, y = y, z = z, \ldots \).

From equation (6), we have,

\[
\frac{1}{n-1} \sum_{i=1}^{n} (\Delta R_i)^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[ \left( \frac{\partial R}{\partial x} \right)^2 (\Delta x_i)^2 + \left( \frac{\partial R}{\partial y} \right)^2 (\Delta y_i)^2 \\
+ \left( \frac{\partial R}{\partial z} \right)^2 (\Delta z_i)^2 + 2 \left( \frac{\partial R}{\partial x} \right) \left( \frac{\partial R}{\partial y} \right) \Delta x_i \Delta y_i \\
+ 2 \left( \frac{\partial R}{\partial x} \right) \left( \frac{\partial R}{\partial z} \right) \Delta x_i \Delta z_i + 2 \left( \frac{\partial R}{\partial y} \right) \left( \frac{\partial R}{\partial z} \right) \Delta y_i \Delta z_i \\
+ \ldots \right] \tag{7}
\]

Since \( x, y, z, \ldots \) are assumed to be independent, the terms \( \Delta x_i \Delta y_j, \Delta x_i \Delta z_j, \Delta y_i \Delta z_j, \ldots \) are all as likely to be positive as to be negative, with the result that they almost entirely
cancel out each other within each of the summations $\sum_i \Delta x_i \Delta y_i$, $\sum_i \Delta x_i \Delta z_i$, $\sum_i \Delta y_i \Delta z_i$, ... provided the number of terms $n$ is sufficiently large.

Thus,

$$\sigma_R^2 = \left( \frac{\partial R}{\partial x} \right)^2 \sigma_x^2 + \left( \frac{\partial R}{\partial y} \right)^2 \sigma_y^2 + \left( \frac{\partial R}{\partial z} \right)^2 \sigma_z^2 + \ldots \quad (8)$$

Equation (8) shows us how the errors occurring in the primary measurements get carried into the final result. This is the proper expression to use for establishing the accuracy of the final result.

6. **COMMON SENSE TREATMENT OF EXPERIMENTAL ERROR**

In elementary practical classes, usually only a few readings are taken for a quantity measured and quite often only one reading is taken. Under the circumstances, a full treatment of error analysis by statistical method is both unnecessary and meaningless. Equations (2), (4), (5) and (8) will be replaced by simpler formulae which are based on a common sense approach. The following points are recommended:

(a) **Error of a measured value:**

The greatest care must be taken in the reading of a value. Normally, readings should be taken to a fraction of the smallest division of the scale. If possible **three readings at least** should be taken of a single quantity to make sure that they are consistent and to obtain some idea of the spread in the readings. If they are not consistent to a sufficient degree, more readings should be taken.

If several readings are available, their mean is taken to be the best value for the measured quantity. Since the number of readings is small, the mean deviation of a single reading (rather than the standard deviation of the mean)
\[ \eta = \frac{1}{n-1} \sum_{i=1}^{n} |x_i - \bar{x}| \]

may be taken to be the error in the mean. The standard deviation of the mean, or the standard error, may be used as the error if \( n \geq 10 \).

Normally, all readings taken should be used in evaluating the mean. However if one of the readings appears to be out of place, it may sometimes be considered a blunder and discarded. For example, in the set of readings:

\[ 10.1, 10.0, 10.0, 10.2, 12.3, 10.1, 10.1, 10.0, 10.1, 10.2 \]

the mean and the mean deviation of the readings from the mean calculated excluding 12.3 are respectively 10.1 and 0.06. The difference 12.3 - 10.1 = 2.2 is 37 times the mean deviation. The probability of the reading 12.3 being a result of large random error is thus almost nonexistent. It appears that some mistake was made in recording this reading and it is reasonable to reject on these grounds.

If only one or two readings are available or if the necessary instrument used is so coarse that the readings are all the same, the error should be estimated on the basis of instrument precision or sensitivity. A rule of thumb is that \( \frac{1}{4} \) to \( \frac{1}{2} \) of the smallest scale is taken to be the error.

If systematic errors are suspected and cannot be avoided or corrected for, e.g., the ruler cannot be brought to be in contact with the measured specimen, the position of the centre of a lens cannot be ascertained precisely etc., some allowance should be made for it in estimating the error of the measured quantity.

In a bridge or potentiometer experiment, we should investigate how far away from the balance point we can move the jockey without producing an
appreciable galvanometer deflection. This should be taken as an estimate of the random error in the balance position.

(b) **Propagation of errors**

Suppose the value of a quantity $R(x, y, z, \ldots)$ is determined from the measured values of a number of independent quantities $x, y, z, \ldots$ which are directly measured. Let $e_x, e_y, e_z, \ldots$ be the errors associated with the measurements of $x, y, z, \ldots$, respectively.

The estimated maximum error in $R$ is given by

$$e_R = \left| \frac{\partial R}{\partial x} \right| e_x + \left| \frac{\partial R}{\partial y} \right| e_y + \left| \frac{\partial R}{\partial z} \right| e_z + \ldots$$

where all the partial deviations are evaluated at $x = \bar{x}, y = \bar{y}, z = \bar{z}, \ldots$. The value of $e_R$ may be taken to be the error in $\langle R \rangle = R(x, y, z, \ldots)$, the measured value of $R$.

As example it is easy to show that:

(i) If $R = ax + by + \ldots$, then $e_R = |a| e_x + |b| e_y + \ldots$

The error in the sum or difference of two or more quantities is the sum of their errors.

(ii) If $R = x^p y^q \ldots$, then $e_R = x^p \left( \frac{e_x}{x} \right) + y^q \left( \frac{e_y}{y} \right) + \ldots$

The percentage error in the product or quotient of two or more quantities is the sum of their percentage errors.
Note that the maximum errors are always larger than those calculated using the statistically correct formula, equation (8).

7. **SIGNIFICANT FIGURES**

The statement that the diameter of a metal sphere is 5.6 cm has a different meaning from the statement that its diameter is 5.60 cm, for the first statement implies that the value is probably reliable to a few times 0.1 cm only, while the latter implies that the value given is probably reliable to a few times 0.01 cm and is certainly reliable to 0.1 cm. The number of digits necessary to express a measurement so as to give some indication of its accuracy is called the number of significant figures. A common rule for deciding the number of significant digits to use is that the accuracy implied should be consistent with that implied by its standard deviation or error. Thus the result should be expressed to the same decimal place as its standard deviation or error while the latter is usually expressed to one or two significant figures.

For elementary practical classes, it is generally sufficient to express errors to one significant figure only. This means that rough calculation is sufficient for error analysis. At each stage, two figures only need to be retained, and the errors of small corrections that have to be applied to the main factor can be disregarded. For example the focal length of a lens is found to be 20.46 cm with an error of 1%, it should be written as 20.5 ± 0.2 cm.

8. **PLANNING OF AN EXPERIMENT**

The accuracy in the final result is determined by the accuracy of the least accurate factor in its expression. It follows that it is meaningless to measure a quantity to high accuracy if some of the quantities involved cannot be accurately measured. For example, in the determination of Young's modulus, $Y$

$$Y = \frac{F}{\frac{\partial L}{\partial A}} = \frac{Mg}{\frac{\partial L}{\partial \pi r^2}}$$
The mass $M$ of the stretching weight can be accurately measured. However, $\delta L$, which is the difference of two measurements, usually has a relatively large uncertainty. Suppose we have
\[
\delta L \approx 0.2 \text{ mm}
\]
and
\[
\frac{\Delta(\delta L)}{\delta L} \approx \frac{0.01}{0.2} \approx 0.05 = 5\%
\]

Then it is not worthwhile to measure $M$ to more accurate than 0.5%. If $M = 2.5$ kg, $\Delta M = 2.5 \times 10^3 \times 0.5 \times 10^2 = 12.5$ gm is sufficiently accurate. On the other hand greatest care should be exercised in observing $L$. Also, as the maximum percentage error of $Y$ is
\[
\frac{\Delta Y}{Y} \approx \left| \frac{\Delta M}{M} \right| + \left| \frac{\Delta(\delta L)}{\delta L} \right| + \left| \frac{\Delta L}{L} \right| + \left| \frac{2\Delta r}{r} \right|
\]
in which the percentage error of $r$ has a factor of two. It should therefore be measured with great care. Actually, if $r = 1$ mm,
\[
\frac{\Delta r}{r} \approx \frac{0.002 \text{ mm}}{1 \text{ mm}} = 0.2\%
\]
is the best accuracy one can obtain with a micrometer. If we are careless and read only to 0.01 mm, then as
\[
\frac{2\Delta r}{r} = 2\%
\]
a 2% additional error will be introduced in the value of $Y$.

In planning an experiment we should have in view the final overall accuracy required. It should be planned in such a way that this aim may be achieved with the minimum effort. The same effort applied to the measurement of different quantities may not have the same effect on the overall accuracy. The relative accuracy to be achieved for each measured quantity must be considered before beginning an experiment.
9. **DISCREPANCY FROM A STANDARD RESULT**

Too frequently, the student compares his result with the standard value from a table and takes the difference as his experimental error. Actually the difference is a discrepancy and not the experimental error. The specimen measured may be different from that which is referred to in the table, or the measuring conditions may be different. In any case, even if the specimen and the conditions are the same, without a proper estimation of the experimental error it is impossible to know if the discrepancy is the result of random errors, or is due to some unrecognized systematic error inherent in the experiment, or if the theory underlying the experiment is faulty. When a quantity is determined using two or more independent methods, it is important to see if the results agree within the experimental accuracy. If the discrepancy cannot be accounted for by random errors, a search for its cause may yield significant physical information.

At the conclusion of a determination, the statement to be made is whether the value obtained agrees with the standard value given in tables within the experimental errors. If this is not so an attempt should be made to account for the discrepancy.

10. **LEAST SQUARES FIT TO A STRAIGHT LINE**

In some experiments, instead of making a number of measurements of a single quantity $x$, we make a series of $N$ measurements of the pair of quantities $(x_i, y_i)$, for $i$ runs from 1 to $N$. We often wish to determine the relationship of the quantity, $y$ as a function of the other quantity $x$. In this section, we will consider the problem of pairs of variables $(x_i, y_i)$ that are linearly related to one another where the relationship is given as

$$y(x) = a + bx$$  \hspace{1cm} (10)

We shall consider in this section a method for determining the most probable values for the coefficients $a$ and $b$. 
Our data consist of a sample of observations drawn from a parent distribution that determines the probability of making any particular observation. For the particular problem of an expected linear relationship between dependent and independent variables, we define parent parameters $a_o$ and $b_o$ such that the actual relationship between $y$ and $x$ is given by

$$y_o(x) = a_o + b_o x$$  \hspace{1cm} (11)

Assume each individual measured value of $y_i$ is itself drawn from a Gaussian distribution with mean $y_o(x_i)$ and standard deviation $\sigma$. The probability $P_i$ for making the observed measurement $y_i$ with standard deviation $\sigma_i$ for the observations about the actual value $y_o(x_i)$ is given by

$$P_i = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-(y_i - y_o(x_i))^2 / 2\sigma_i^2}$$  \hspace{1cm} (12)

The probability for making the observed set of measurements of the $N$ values of $y_i$ is the product of the probabilities for each observation:

$$P(a_o, b_o) = \prod \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\Sigma[(y_i - y_o(x_i))^2 / 2\sigma_i^2]}$$  \hspace{1cm} (13)

where the product is taken with $i$ ranging from 1 to $N$.

Similarly, for any estimated values of the parameters $a$ and $b$, we can calculate the probability of obtaining the observed set of measurements

$$P(a, b) = \prod \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\Sigma[(y_i - y(x_i))^2 / 2\sigma_i^2]}$$  \hspace{1cm} (14)

where $y(x_i) = a + bx_i$. Since we assume that the observed set of measurements is coming from the parent distribution of equation (13), the probability of equation (13) is the maximum probability attainable with equation (14). Thus, the maximum likelihood
estimates for a and b are those values that maximize the probability of equation (14). This is equivalent to minimizing the sum in the exponential.

Define this sum as the goodness-of-fit parameters:

$$\chi^2 = \sum \left[ \frac{y_i - y(x_i)}{\sigma_i} \right]^2 = \sum \left[ \frac{1}{\sigma_i} (y_i - a - bx_i) \right]^2$$  \hspace{1cm} (15)

To find the optimum fit to the data will be to find values of a and b that minimize this sum. This is known as the method of least square fit.

To find the values of the parameters a and b that yield the minimum values for $\chi^2$, the partial derivatives of $\chi^2$ w.r.t. a and b are set to zero

$$\frac{\partial}{\partial a} \chi^2 = \frac{\partial}{\partial a} \sum \left[ \frac{1}{\sigma_i^2} (y_i - a - bx_i) \right]^2 = -2 \sum \left[ \frac{1}{\sigma_i^2} (y_i - a - bx_i) \right] = 0$$

$$\frac{\partial}{\partial b} \chi^2 = \frac{\partial}{\partial b} \sum \left[ \frac{1}{\sigma_i^2} (y_i - a - bx_i) \right]^2 = -2 \sum \left[ \frac{1}{\sigma_i^2} (y_i - a - bx_i)x_i \right] = 0$$

These equations can be rearranged as a pair of linear simultaneous equations in the unknown parameters a and b:

$$\sum \frac{y_i}{\sigma_i^2} = a \sum \frac{1}{\sigma_i^2} + b \sum \frac{x_i}{\sigma_i^2} \hspace{1cm} (16)$$

$$\sum \frac{x_i y_i}{\sigma_i^2} = a \sum \frac{x_i}{\sigma_i^2} + b \sum \frac{x_i^2}{\sigma_i^2} \hspace{1cm} (17)$$
The solutions are given by

\[ a = \frac{1}{\Delta} \left( \sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} \right) \]  

(18)

\[ b = \frac{1}{\Delta} \left( \sum \frac{1}{\sigma_i^2} \sum x_i y_i - \sum \frac{x_i}{\sigma_i^2} \sum y_i \right) \]  

(19)

\[ \Delta = \sum \frac{1}{\sigma_i^2} \sum x_i^2 - \left( \sum \frac{x_i}{\sigma_i^2} \right)^2 \]  

(20)

Note that in the special case that all the uncertainties are equal (\( \sigma = \sigma_i \)), they cancel and the solutions may be written as

\[ a = \frac{1}{\Delta'} \left( \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \right) \]  

(21)

\[ b = \frac{1}{\Delta'} \left( N \sum x_i y_i - \sum x_i \sum y_i \right) \]  

(22)

\[ \Delta' = N \sum x_i^2 - \left( \sum x_i \right)^2 \]  

(23)

Uncertainties in the parameters, to determine the uncertainties in the parameters \( a \) and \( b \), we take the partial derivatives of equations above (for \( a \) and \( b \))

\[ \sigma_a^2 = \frac{1}{\Delta} \sum \frac{x_i^2}{\sigma_i^2} \]  

(24)

\[ \sigma_b^2 = \frac{1}{\Delta} \sum \frac{1}{\sigma_i^2} \]  

(25)
Again for special case of the uncertainties being the same

\[ \sigma_a^2 = \frac{\sigma^2}{\Delta'} \sum x_i^2 \]  
(26)

\[ \sigma_b^2 = N \frac{\sigma^2}{\Delta'} \]  
(27)

\[ \sigma^2 = \frac{1}{N-2} \sum (y_i - a - bx_i)^2 \]  
(28)

For least square fit to a polynomial and other arbitrary function, students are strongly encouraged to refer to the references.

11. REFERENCES

Philip R. Bevington and D. Keith Robinson : Data Reduction and Error Analysis for the Physical Sciences
H.F. Meiners, W. Eppenstein, K.H. Moore : Laboratory Physics
R.M. Whittle, J. Yarwood : Experimental Physics for Students
J. Topping : Errors of Observation and Their Treatment
Y. Beers : Introduction to The Theory of Errors
G.L. Squires : Practical Physics
EXERCISES 1

1. In the measurement of a quantity \( x \), the following distribution is obtained, where \( f \) is the frequency with which a given value of \( x \) appears. If the distribution may be assumed to be normal, find the mean, the standard deviation and the standard error of \( x \). Compare the values of the standard deviation obtained by using eq. (2) and eq. (4) separately.

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c}
 x & 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 & 29 \\
 f & 1 & 5 & 17 & 49 & 85 & 52 & 25 & 11 & 4 & 1 \\
\end{array}
\]

2. The refractive index \( \mu \) of the glass of a prism for a particular wavelength is obtained from the angle of minimum deviation \( D \),

\[
\mu = \frac{\sin \left(\frac{1}{2}(A + D)\right)}{\sin \left(\frac{1}{2}A\right)}
\]

where \( A \) is the apex angle of the prism.

(a) Show that the maximum error in \( \mu \) is given by

\[
e_{\mu} = \left| \sin \left(\frac{1}{2}D\right) \right| e_A + \left| \cos \left(\frac{1}{2}(A + D)\right) \right| e_D
\]

where \( e_A \) and \( e_D \) are in radians.

(b) Three measurements of \( A \) and \( D \) gave

\[
\begin{align*}
A &= 60°5.4', 60°4.8', 60°5.0'; \\
D &= 46°37.0', 46°36.2', 46°36.6'.
\end{align*}
\]

Find \( \mu \) and its error.
3. In the determination of \( g \) by means of a simple pendulum, the following data are obtained:

\[ \ell = 99.4, 99.5, 99.3 \text{ cm.} \]

Time for 20 oscillations = 40.0, 40.0, 39.8, 40.2, 39.9 sec.

(a) Using \( T = 2\pi \sqrt{\ell/g} \), show that

\[
\frac{e_g}{\langle g \rangle} = \frac{e_\ell}{\ell} + \frac{2e_T}{T}
\]

where \( \langle g \rangle = 4\pi^2 \frac{\ell}{T^2} \).

(b) Calculate \( g \) and its error.

4(a) The focal length of a thin converging lens may be determined by measuring the object and image distances from the lens and using the formula

\[
\frac{1}{f} = \frac{1}{u} + \frac{1}{v}
\]

Show that

\[
e_f = \langle f \rangle^2 \left( \frac{e_u}{u^2} + \frac{e_v}{v^2} \right)
\]

where \( \langle f \rangle = \frac{uv}{u+v} \).

(b) The object and image distances from a thin converging lens are \( u = 15.2 \pm 0.3 \) cm, \( v = 6.4 \pm 0.2 \) cm. Find the focal length of the lens and its error.
5. A student measures the following pairs of quantities for \((x_i, y_i)\) during the experiments. Assuming that the relation between the quantities is linear \(y_i = a + bx_i\) and assume uniform uncertainties \(\sigma = \sigma_i\) in \(y_i\), find the values of the coefficients \(a\) and \(b\) and the uncertainties in \(a\) and \(b\).

<table>
<thead>
<tr>
<th>(x_i)</th>
<th>(y_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>102</td>
</tr>
<tr>
<td>0.2</td>
<td>158</td>
</tr>
<tr>
<td>0.3</td>
<td>209</td>
</tr>
<tr>
<td>0.4</td>
<td>268</td>
</tr>
<tr>
<td>0.5</td>
<td>298</td>
</tr>
<tr>
<td>0.6</td>
<td>359</td>
</tr>
<tr>
<td>0.7</td>
<td>411</td>
</tr>
<tr>
<td>0.8</td>
<td>461</td>
</tr>
<tr>
<td>0.9</td>
<td>505</td>
</tr>
<tr>
<td>1.0</td>
<td>558</td>
</tr>
</tbody>
</table>

Note that in this question, you will be using the PC facilities in the department and learn how to make use of the EXCEL software for the computation. The lab instructor will go through the details on how to make use of the software to carry out the calculation during the lab class.
6. A student makes use of a thermal technique to synthesize Copper Oxide nanowires onto top of Zinc Oxide nanowires in the laboratory. The following table summarizes the total length of the nanowires as a function of the growth time:

<table>
<thead>
<tr>
<th>Growth Time (Hours)</th>
<th>Length of Nanowires (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1400</td>
</tr>
<tr>
<td>1</td>
<td>1100</td>
</tr>
<tr>
<td>1</td>
<td>1270</td>
</tr>
<tr>
<td>2</td>
<td>1710</td>
</tr>
<tr>
<td>2</td>
<td>1530</td>
</tr>
<tr>
<td>2</td>
<td>1680</td>
</tr>
<tr>
<td>3</td>
<td>1850</td>
</tr>
<tr>
<td>4</td>
<td>2250</td>
</tr>
<tr>
<td>5</td>
<td>2520</td>
</tr>
<tr>
<td>5</td>
<td>2590</td>
</tr>
<tr>
<td>5</td>
<td>2480</td>
</tr>
<tr>
<td>7</td>
<td>3050</td>
</tr>
<tr>
<td>7</td>
<td>3140</td>
</tr>
<tr>
<td>8</td>
<td>3250</td>
</tr>
<tr>
<td>9</td>
<td>3800</td>
</tr>
<tr>
<td>10</td>
<td>4100</td>
</tr>
<tr>
<td>15</td>
<td>5600</td>
</tr>
<tr>
<td>15</td>
<td>5410</td>
</tr>
<tr>
<td>15</td>
<td>5380</td>
</tr>
<tr>
<td>20</td>
<td>6800</td>
</tr>
<tr>
<td>20</td>
<td>7100</td>
</tr>
<tr>
<td>20</td>
<td>6900</td>
</tr>
<tr>
<td>24</td>
<td>8300</td>
</tr>
<tr>
<td>24</td>
<td>8400</td>
</tr>
<tr>
<td>24</td>
<td>8150</td>
</tr>
<tr>
<td>24</td>
<td>8000</td>
</tr>
</tbody>
</table>

(a) To test whether or not the length of the nanowires varies linearly with the growth time, plot a graph of the length of the nanowires versus the growth time.

(b) Determine the rate of growth of the copper oxide nanowires. What is the uncertainty in the rate of growth?

(c) Determine the length of the Zinc Oxide nanowires. What is the uncertainty in the length of the Zinc Oxide Nanowires?