Principle of Data Analysis
Supplementary Material for Physics Laboratory I
Physics Department
The Chinese University of Hong Kong
(Revised June 1998)

Contents

1 Error Estimation .............................................................................................................. 2
  1.1 Sources of Errors ........................................................................................................ 2
  1.2 Scale Uncertainties ..................................................................................................... 2
  1.3 Statistical Errors ......................................................................................................... 3
  1.4 Systematic Errors ....................................................................................................... 6
  1.5 Interpretation of the Experiment Error .................................................................... 6
  1.6 Calculation of Mean and Standard Error by using Excel ........................................ 6

2 Graphical Analysis ............................................................................................................ 8
  2.1 The Linear Equation ................................................................................................. 8
  2.2 Conversion to Linear Graphs .................................................................................... 9
1 Error Estimation

Science depends on experimental observation and measurement. However, the measured values have no useful physical significance unless we specify the measuring processes and their uncertainties or errors. By error we do not mean that a mistake has been made. Error is an estimate of the accuracy of the measured result. Frequently, error estimation of the experiment result has to be made before drawing any conclusion.

1.1 Sources of Errors

According to the natures of the errors, they can be divided into the following three types.

1. **Scale uncertainties** represent the smallest quantities that can be estimated on the scale of instruments.
2. **Statistical errors** are fluctuations in the measurements due to random nature.
3. **Systematic errors** are definite falsifications of the measured quantity, making it either larger or smaller than it should be.

**Plain mistake** such as a misreading of a scale is not included in this list — such errors are preventable by careful work.

1.2 Scale Uncertainties

The scale uncertainty of a measurement depends on the nature and setting of the instrument. As an example, let’s consider measuring the voltage across a resistor with a digital multimeter. There are a few different settings for a digital multimeter to measure voltage. Each setting has a certain accuracy and an upper limit of voltage that it can measure. Assume that the voltage across the resistor is known to be 2.3456 V. The reading of the multimeter will give 2.3 V if a full range of 200V is chosen. This reading cannot differentiate voltage between 2.25 V and 2.35 V. This means that the accuracy of the reading is only 0.05 V and this is not very satisfactory. To improve the accuracy, we can set the full range to 20 V and the reading will be 2.35 V and the uncertainty is now 0.005 V. If we attempt to further improve the accuracy by setting the full range to 2 V, the meter will give an overflow signal. Therefore, the measured voltage is 2.35 ± 0.005 V instead of the expected value of 2.3456 V. If a more accurate result is needed, you would need a digital multimeter (or voltmeter) with better performance.

![Digital multimeter](image_url)

Fig. 1 Voltage measurement with a digital multimeter.

From the above example, we see that the accuracy of a digital instrument cannot be better than half of its least significant digit. While the accuracy of a digital instrument is quite well determined, it is not so obvious for an analog instrument. If the same voltage is measured with an analog voltmeter (Fig. 2), there might be a few different conclusions for the measurement.
Depending on how careful the observer takes the reading, the result of the measurement can be
2.3 ± 0.05 V or 2.35 ± 0.02 V. The accuracy of the result can be anywhere from one-half to one-fifth of
the smallest division.

1.3 Statistical Errors

If an experimental result is calculated from the mean of many measurements, its uncertainty can then be
estimated from the distribution of the measured results.

Example 1:

Let’s measure the resistance $R$ of a resistor by a voltmeter and an ammeter.

The accuracy of individual measurement is about

$$\delta R \approx \frac{0.05}{1.5} = 3.3\%.$$
which is only
\[
\frac{\delta R}{R} = \frac{1.226 - 1.234}{1.234} = -0.6\%
\]
off from the expected value. In fact, the accuracy of the experiment result can be further improved if more measurements are taken.

1.3.1 Uncertainty of a single measurement due to scale uncertainty

If experimental result is calculated from a single measurement, its error can be usually estimated from the scale uncertainties of the instruments. In Example 1, the resistance \( R \) is calculated from the measurements of voltage \( V \) and current \( I \). The error of a single \( R \) value is then due to the scale uncertainties of the voltmeter and ammeter. Since the scale uncertainty of the voltmeter is negligible, the error of \( R \) is due solely to the scale uncertainty of the ammeter which is 0.05 mA. For multiplication and division, it is better to handle the error in percentage form. In this example, the percentage error of the current \( I \) is roughly \( 0.05/1 = 5\% \). Therefore, the percentage error of the resistance \( R \) should also be 5%.

For simplicity, students only need to consider the largest source of error during calculations in Physics Laboratory I. Consider the one with largest percentage error for multiplication or division and consider the one with largest absolute error for addition or subtraction.

**Exercise 1:**

The resistance \( R_1 \) of a resistor is measured by voltage-current method. The voltage \( V_1 \) measured is 1.2 ± 0.05 V and the current \( I_1 \) measured is 2.4 ± 0.05 mA.

i. Find the resistance \( R_1 \) and its error.

ii. What is the total resistance \( R \) and its error if the resistor \( R_1 \) is in series with an resistor \( R_2 = 1.0 \pm 0.05 \) kΩ?

1.3.2 Significant figures

In carrying out computation, only significant figures should be retained. Significant figures are those figures which are known to be reasonably trustworthy. In Example 1, the result of \( R \) should be given by at least three figures since the error is as high as 10% if only two figures are kept. Moreover, more than three figures should not be used since it will imply an accuracy better than it really is (e.g. four figures imply an accuracy of about 0.1%).

The number of figures of the result should be kept only up to the same digit as the error of the result.

**Exercise 2:**

The experimental result and measured error of Example 1 depend on the number of significant figures used for the calculation of the resistance \( R \) for each of the measurements.

i. Repeat the calculation for \( \bar{R} \) and \( \delta \bar{R} \) by keeping only two significant figures for \( R \).

ii. Repeat the calculation for \( \bar{R} \) and \( \delta \bar{R} \) by keeping four significant figures for \( R \).

How many significant figures should be used for the calculation of \( R \)?

1.3.3 Standard deviation

If many measurements \( x_i \) are taken for a quantity, it is possible to estimate the error \( \delta x \) of a single measurement from the distribution. If the exact value \( a \) of the quantity is already known, the error of a single measurement can be defined as the root-mean-square of the deviation from the exact value.
\[ \delta x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \]

However, this is usually not the case for a real experiment (otherwise, there is no need to perform the experiment). Instead of the exact value, the mean \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) of the measurement can be used in the estimation of error

\[ \delta x = s \equiv \frac{1}{n-1} \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \]

where \( s \) is called the **standard deviation** of the set of measurements. Note that \( n-1 \) is used in the denominator instead of \( n \) since the standard deviation is undefined for \( n = 1 \).

**Exercise 3:**

Calculate the standard deviation of the set of measurements in Example 1 and compare it with the root-mean-square of the deviations from the exact value.

Compare it also with the one calculated from the scale uncertainty of the ammeter.

### 1.3.4 Standard error

When the experimental result is taken as the mean of a set of measurements, its error is much smaller than the one of individual measurement. The error of a mean is called the standard error which is defined as

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

**Exercise 4:**

The resistance per unit length \( r \) of a metal wire is determined by measuring the resistance \( R \) of a single wire. The following results are obtained.

<table>
<thead>
<tr>
<th>Length of wire ( l )</th>
<th>Voltage ( V ) (V) ± 0.05</th>
<th>Current ( I ) (mA) ± 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 ± 0.05</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

i. Calculate the resistance \( R \) for each measurement.
ii. Determine $\bar{R}$ and $\delta R$.

iii. Calculate the resistance per unit length $\rho$ for each measurement.

iv. Determine $\bar{\rho}$ and $\delta \rho$.

v. What should be the experiment result for $\rho$ and $\delta \rho$?

1.4 Systematic Errors

Systematic errors are definite falsifications of the measured quantity owing to factors which have not been properly taken into account. While it is relatively easy to handle scale uncertainties and statistical errors, it is very hard to determine systematic errors. There is no standard procedure to calculate or estimate the systematic errors, but most of them can be avoided or discovered if proper experimental techniques are used. The following lists some common experimental techniques that can help to minimize the effect of systematic errors.

1. Calibration should be done before using any instrument. This can be done by checking the zero point of the instrument and taking one or a few measurements with a standard reference source.

2. Compare the error of a single measurement calculated from the scale uncertainties with the standard deviation of the measured data. The standard deviation should be smaller than the error calculated from the scale uncertainties. If this is not the case, there is systematic error which is comparable or even more important than the scale uncertainties in the data.

3. Compare the experiment result with the result from another independent experiment. If the results disagree, at least one of the experiment results has systematic error.

In Physics Laboratory I, students should check the zeroing of the instruments before they are used. Students should also compare the result with the accepted or given value and discuss the sources of systematic errors.

1.5 Interpretation of the Experiment Error

Usually, a theory is considered to be verified when the experiment result is within one standard error of the calculated value. On the other hand, it is considered to be disproved when the experiment result is more than three standard errors away from the calculated value. Of course, careful checking of the experiment result and search for systematic error should be carried out before drawing such a conclusion. When the experiment result is about two standard errors from the calculated value, the experiment itself should then be improved or more data should be collected before drawing any conclusion.

In Physics Laboratory I, students will only study proved theories or measured quantities. Derivations from the expected values will only imply that systematic errors exist in the experiments.

Students should look for sources of systematic errors and approaches to improve the experiments no matter what are the differences between the experiment results and the expected values.

1.6 Calculation of Mean and Standard Error by using Excel

While it is simple and straightforward to calculate the mean and standard error for a few data points, it becomes tiresome for a large data set. To avoid unnecessary mistakes, students are advised to use calculators or computers to perform the calculations. Since Excel is one of the most popular software, it will be helpful to use Excel for some basic error calculations.

In the following sections, we will introduce how to use Excel to calculate the mean and standard error for a set of data.
1.6.1 Using Data Analysis Tools

1. Open Excel and enter the data.
2. Select Data Analysis ( altında $\text{Data Analysis}$ $\text{Tools}$ $\text{Descriptive Statistics}$ $\text{Input Range}$ $\text{Output Range}$ $\text{Summary statistics}$ ) in the menu bar. A window for Data Analysis will be displayed.

3. Select Descriptive Statistics $\text{Descriptive Statistics}$ $\text{Input Range}$ $\text{Output Range}$ $\text{Summary statistics}$ from the Data Analysis window. A window for Descriptive Statistics will be displayed.

4. Enter the range of data at the Input Range ( $\text{Input Range}$ $\text{Output Range}$ $\text{Summary statistics}$ ) and click OK ( $\text{OK}$ $\text{Cancel}$ $\text{Help}$ ).

5. The summary of statistics will be listed in a new worksheet.

1.6.2 Using Statistical Functions

1. Open Excel and enter the data.
2. Use the statistical function AVERAGE(A1:An) to calculate the mean $\bar{x}$. 
3. Use the statistical function STDEV(A1:An) to calculate the standard deviation $s$.
4. The standard error $\sigma$ can be calculated from the standard deviation $s$, $\sigma = s/\sqrt{n}$ where $n$ is the number of data entries.
2 Graphical Analysis

In the modern experiment, most of the data analyses are performed by computer. However, it is still useful to study the relations among the measurements by looking at some simple plots of the data. Very often, obvious mistakes can be quickly discovered from the graphs in the early stage of the experiment and modifications of the experiment can be done before a lot of time has been wasted.

In this section, you will learn some basic graphing techniques and methods to extract information from a graph.

In Physics Laboratory I, students are asked to plot the graph by hand instead of using computer.

2.1 The Linear Equation

When a linear equation \( y = ax + b \) is plotted, a straight line is obtained. For two points \((x_1, y_1)\) and \((x_2, y_2)\) on the straight line, the slope \( a \) is defined as

\[
a = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}
\]

The value of \( y \) when \( x = 0 \) is \( b \), and called the y-intercept.

2.1.1 Data Fitting

Suppose you have performed an experiment to investigate the relationship between two variables \( x \) and \( y \). The resulting plot is shown below.

![Graph of a linear equation](image)

Judging from the graph you probably conclude that the relationship is linear and you can simply draw a straight line \( y = ax + b \) by eye. The uncertainties of the results \( \delta a \) and \( \delta b \) can be estimated to be the difference between the best (solid) line and an alternate (dash) line.

This data fitting technique is required in Physics Laboratory I. More exact analysis methods will be introduced in advanced laboratory courses.

2.1.2 Labeling Graphs

Every graph should be identified with a title, e.g. “Acceleration of a glider on an inclined air track”, not just “acceleration”. Label each axis, including the units used, e.g. “Acceleration in cm/sec\(^2\)” is acceptable while just using the symbol “\( a \)” is meaningless.
2.1.3 What Units to Plot In

Sometimes measurements are taken in unconventional units, e.g. your measurement may be in thousandths of an inch and in fifteenths of a second, but you are interested in finding the acceleration \( a \) in \( \text{cm/sec}^2 \) from the slope of a \( v-t \) plot. Clearly, you can first convert each distance to cm and then divide each distance by 1/15 sec to obtain a velocity. If you have 25 points, you have 25 \( \times \) 2 chances for making mistake. It might be better to plot the raw data. You can then find \( a \) in unconventional units and can make a single conversion to \( \text{cm/sec}^2 \).

**Exercise 5:**

The distance traveled, \( \Delta x \), in each time interval \( \Delta t = 1/15 \) sec is given below, in units of \( 10^{-3} \) inch.

\[
\Delta x = 0.12, 0.31, 0.49, 0.71, 0.88
\]

Plot \( \Delta x \) versus time, and find the slope and its error. This is essentially a velocity-time graph since \( \Delta x \propto v \). Hence from the slope deduce the acceleration and its error in \( \text{cm/sec}^2 \). Pay attention to units at each step.

2.2 Conversion to Linear Graphs

When an exponential equation \( y = ae^{bx} \) is plotted, a curve is obtained. However, there is no simply method to extract the parameters \( a \) and \( b \) from the curve. In order to extract useful information, straight line should be plotted instead of a curve. Fortunately, the exponential equation can be easily converted into a linear equation.

\[
y = ae^{bx} \\
\Rightarrow \ln y = \ln a + bx
\]

The parameters of an exponential equation can then be determined if the logarithm of the data are plotted.

**Exercise 6:**

The velocity \( v \) of a ball falling through oil is expected to change with time \( t \) as \( v = v_0e^{-t/T} \).

<table>
<thead>
<tr>
<th>( t ) (sec)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v ) (cm sec(^{-1}))</td>
<td>29.4</td>
<td>14.5</td>
<td>6.54</td>
<td>3.48</td>
<td>1.72</td>
<td>0.840</td>
<td>0.411</td>
<td>0.203</td>
</tr>
</tbody>
</table>

Plot a suitable graph and deduce the constants \( v_0 \) and \( T \). Show a vertical scale for \( \ln v \) (or \( \log v \)) on the right hand side (this will be equally spaced) and show another vertical scale for \( v \) on the left hand side (this will not be equally spaced).

Other functional relations can often be converted to a linear relation, as in the following example.

**Exercise 7:**

The specific heat \( C \) of a material is measured at various absolute temperatures \( T \). It is believed that \( C \) and \( T \) are related as

\[
C = aT^b + bT
\]

where \( a, b \) are constants. What should be plotted in order to obtain a linear graph? State how \( a \) and \( b \) can be obtained from the slope and the intercept.