Exercise in Statistics

Before you start any experiments it is important that you have a good understanding of techniques of error analysis. Appendix A contains a summary of what you need to know, you should read through this document before proceeding with this lab. The exercises contained in this lab will give you a chance to apply these methods.

Exercise 1

1. Take ten toothpicks. The last page of this document has two lines drawn with a spacing equal in length to a toothpick. Place the page onto a flat surface. Hold the ten toothpicks about 30 cm(1 foot) above the page, centered between the lines, and drop them so that they scatter upon the page. Some of the toothpicks will lie wholly within the lines; some will lie outside the lines, and others will cross a line. In this exercise we will take statistics of those that cross a line. Count the number, n, of toothpicks that cross a line and record the value, this is the 1st trial. Repeat for a total of 100 trials. The table below is a convenience to record your data into.

Trial #	1st Ten trials	2nd Ten trials	3rd ten trials	4th ten trials	5th ten trials	6th ten trials	7th ten trials	8th ten trials	9th ten trials	10th ten trials
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
Average										

2. With the measurements now complete, the results can be put into a more visual form for easier comprehension of the data. A bar chart (or histogram) showing the frequency f(n) with which different values of *n* occurs would be ideal for this.

3. A quantitative way of presenting these results is by calculating certain quantities which tell us certain information about the distribution of the results. In most experimental situations only two numbers are really necessary. The first is some measure of the most likely, or average, value of *n* and the second is some measure of the spread in values in *n*. There are several possible definitions we could use for the average value; *mode* (the value which occurs with the greatest frequency); the *median* (when *n* is odd the middle value in an ordered set, when *n* is even the average of the middle two values) and the *mean* of *arithmetic average* defined by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Evaluate the mode, median and mean for your data set.

4. Likewise there are several possible measures of the spread of your measurements. One that comes to mind is the *range* from the highest to lowest value recorded. An alternative is the average deviation from the mean value (which would require taking the modulus of each value before averaging). The most useful is the *variance* or its square root, the *standard deviation*, defined as

$$s \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x-\bar{x})^2}$$

5. Using simple trigonometry, is is possible to show that the probability that an arbitrary toothpick, thrown at random, will fall across a line is

$$\frac{2}{\pi} \approx 0.6366$$

Therefore, in a throw of 10 toothpicks we would expect that an average of 6.366 will cross a line. There will of course, be a spread about this number and it may be shown that the theoretical spread is 2.313. It is highly unlikely that the mean and variance of your measurements are exactly equal to these theoretical values since the number of trials is insufficient to eradicate all the statistical fluctuations that may occur and there may be faults in your experimental technique.

6. To understand how these fluctuations depend on the number of trials let us look at each of the ten set of trials. Evaluate and list the 10 mean values. Notice the variance between each sets mean and the mean of all 100 values. This should display what confidence can be taken for the best estimated value of a mean when only a small number of data samples are used. Does the mean value of the ten mean values agree with the mean of the 100 values?

7. Notice that these mean values fluctuate from one member to the next. This fluctuation is a measure of how well we know the overall mean, i.e. the degree of confidence we should place in the mean we calculated from all 100 values. For each set of ten values evaluate the standard deviation. These values are called the standard deviation of the mean, s_m. The smaller this number, the greater the accuracy of the measurement. There is an alternative method to determine s_m. It turns out that the relationship between s and s_m is

$$s_m \equiv \frac{s}{\sqrt{N}}$$

8. When giving a result for any values determined from the methods outlined it is conventional to quote it as

(mean value) ± (standard deviation of the mean)

Note: The significance of the ± is just that we are dealing with a square root which can be positive or negative. It does not imply that a true value lies somewhere in the range of the ± operation. In fact for nearly all experimental situations it turns out that there is a roughly one in three chance that the true value lies outside this region.

Compare your calculated value of the mean and associated standard deviation with theoretical values given in step 5.

9. The standard deviation is often referred to as the experimental error. This is somewhat misleading as it is a measure of the statistical fluctuations in a measurement. Even if your experimental technique were prefect and you had forgotten nothing, you would not expect to get the same answer for the mean value in every repetition of the experiment.

The standard deviation of the mean is calculated by dividing the standard deviation of the data by $N^{1/2}$. this means that the larger the value of N, the smaller the value of s_m. What value of N would be required to reduce s_m to one tenth its present value? One thousandth of its present value?

10. It is, of course, conceivable that your experimental technique may not have been perfect and you may have forgotten something vital or done something foolish in performing the experiment. In such cases you are not going to find that your mean value approaches the theoretical value no matter how large a value of N you choose. This is an example of systematic error and virtually impossible to eliminate entirely.
Compare you results with that of two others in your lab and include it into your lab report. Is the variance of values within statistical expectations?
If not can they be accounted to systematic errors? How was their approach to gathering the data different than your own?

Consider some of the following systematic errors , would your mean value (a) increase, (b) decrease, (c) be unchanged or (d) impossible to say.

The height at which you release the toothpicks increased.
The height at which you released the toothpicks decreased.
The spacing between the lines is less than the length of a toothpick.
The spacing between the lines is greater than the length of a toothpick.
The lines are not parallel.
The toothpicks are of unequal lengths.
Unknown to you part way into the experiment someone took one of the toothpicks and puts it to us as the manufacture intended.

Exercise 2
Graph Plotting

While graphing by hand using pencil and paper is a tried and true form of displaying information graphically the modern age has computers and software which makes this a practice that is becoming obsolete. The graphs that you will be asked to create in the following sections are to be created utilizing a computer and included within your lab report. Much of the analysis can also be performed using the computer. Some thought may have to be utilized in getting the computer to do your bidding but learning the shortfalls of a program and overcoming them is part of this labs exercise.

Linear Graph.

Using a scale with and accuracy of 1g,and a beaker the following values were obtained as the volume of a liquid within the beaker was increased in 5mL increments.

Vol. (mL)	5	10	15	20	25	30	35	40	45	50	55	60
Mass (g)	14	20	24	28	32	36	42	46	50	54	58	64

Plot a graph of Mass vs volume that includes error bars a best fit line and a worst fit line.
From the graph evaluate the best estimate value of the slope and y-intercept and their uncertainties.

Why is the y-intercept not zero? What quantity does this value represent?
What quantity does the slope represents?

Semi-Log Graph

Very often in science you will face results that have an exponential relation in the form of $y = e^{Cx}$. An example of this is radioactive decay where the number of nuclei decaying in a given interval, N, is related to the time that has elapsed, t, by the equation $N = N_o e^{-Ct}$. To cast this into a suitable straight line form we take the natural logarithms:

$$log_e N = log\ N_e - (C)(t)$$

or to base 10:

$$log_{10} N = log10\ N_o - (0.4343)(C)(t) \text{ which is of the form } y = b - kx$$

(0.4343 being log_{10} e). A plot of $log_{10}N$ vs. t should yield a straight line. For this purpose a semi-log graph is used. The horizontal scale is an ordinary liner scale, but the vertical scale is a logarithmic one, that equal increments of ($log_{10}y$) are provided. The slope of the line being 0.4343 C.

The data given in the table represents the count rate taken over a 30 second interval for a decaying radioactive source, and this measurement was repeated every hour for ten hours.

Radioactive decay											
Time (Hours)	0	1	2	3	4	5	6	7	8	9	10
Count rate (CPS)	800	593	439	325	241	179	132	98	73	54	40

Plot these values of N vs. time using both a regular graph and a semi-log graph.

In times past the semi-log graph was necessary to more easily obtain a value for the constant in the exponent of *e*, by being able to perform a simple slope calculation of $\Delta y / \Delta x$. However, most software programs that graph data also have the ability to fit trend lines to that data, making having to do a semi-log graph more of a choice than a necessity in many cases.
From each graph that you plotted fit a trend line to the graph that will give you a value of C for the equation $N = N_o e^{-Ct}$.

Lastly evaluate from your value of C the half-life of the radioisotope. This is the time for N to decrease by a factor of two.