

# Error Analysis

This brief discussion of error analysis will be adequate for these labs. For a thorough understanding of error analysis and for further information see John R. Taylor, *An Introduction to Error Analysis*, 2nd Ed. (University Science Books, 1997).

## 1 Motivation

Error Analysis is an important tool for all quantitative sciences. Measured values are always subject to uncertainty, whether due to human or machine imperfections. A measured number by itself, without a value for the uncertainty, carries very little meaning as it cannot rule out any other value. For example, a measurement of 5 mm is not inconsistent with one of 20 meters if the uncertainties are large enough. It is therefore important to determine and quote the uncertainty range for every measurement in an experiment, and every relevant result.

## 2 Units

A number is not complete without the units associated with that quantity, and often the uncertainty attached to the number. Be sure to include both for any number given. For example, for a length  $l$

$$l = 2.0 \pm 0.2 \text{ cm} \quad (1)$$

## 3 Significant Figures

The number of significant figures is the number of non-zero digits in the quantity, and number of zero digits between any two non-zero digits. For example  $x = 120.34$  has 5 significant figures. All digits are assumed known except the last digit on the right. This last number contains some error. For  $x = 120.34$  the last digit on the right, 4, might be uncertain to  $\pm 1$ , giving a range for the number of  $x = 120.33$  to  $120.35$ . The uncertainty in this last digit is denoted  $\delta x = 0.01$ , where  $\delta$  is the lowercase greek letter delta. The quantity can be written as  $x \pm \delta x$ .

It is important to consider the significant figures when writing a number. It is similarly wrong to write  $120.3 \pm 0.01$ , since the numbers appearing after 4 are not known within the uncertainty. The number of decimal places in  $\delta x$  should always agree with the number of decimal places in  $x$ .

## 4 Estimating Uncertainty

Often the uncertainty in a measurement depends on the reading on a scale or meter. You must decide how well you can align the pointer or object to the scale. Some examples include

**Meter stick, ruler, etc** If the finest division is 1mm, the uncertainty might be  $\pm 0.5$ mm or perhaps a bit smaller.

**Science Workshop quantities such as length, time, and angle** Use your judgement as to what is the appropriate number of decimal places to use. SWS allows you to change the number of decimal places observed. Consider performing simple experiments to see in which decimal place the measurement becomes uncertain. If you are using the motion sensor place a stationary object in front of it and use the table display to take 10 or 15 measurements of the distance. Note in which decimal place the measurements become uncertain. That is, in which decimal place the numbers deviate significantly.

**Rotary motion sensor** This sensor's window indicated 2 precisions of 360 or 1440 divisions per revolution. This would lead to a **minimum** uncertainty of 1 or  $\frac{1}{4}$  degrees.

**Analog values** SWS used a 12-bit analog-to-digital converter. This means that the voltage interval is broken into  $2^{12} = 4096$  values. If the possible voltage range is from  $-10$  to  $+10$  V the minimum uncertainty would be  $\frac{20}{4096} = 0.005$  V.

**Digital Sensors** SWS usually measures times. The motion sensor, which is digital, measures the time interval between the emission of a short acoustic pulse and the detection of the reflected pulse. The SWS clock has a resolution of .0001 s or .1 ms. Any measurement cannot be better than the uncertainty resulting from this time resolution. There may well be reasons why the uncertainties are larger.

## 5 Random and Systematic Error

**Random Error** A random error is one which is different after repeated measurements, and are distributed randomly around true value. The scatter or uncertainty due to random errors will produce values which are different from measurement to measurement. A special case of random errors are counting measurements. If you are counting something that is truly random, the uncertainty in the number of counts is the square root of the number of counts. That is,

$$\delta\nu = \pm\sqrt{\nu} \quad (2)$$

where  $\nu$  is the number of counts. If you again count for the same amount of time, it is unlikely that you will get the same number of counts. But it is likely that the number of counts will fall within the range  $\nu + \sqrt{\nu}$  to  $\nu - \sqrt{\nu}$ .

**Systematic Error** A systematic error will be the same after repeated measurement and is a consistent deviation from the true value. Incorrect calibration of an instrument will cause systematic error. Experimental circumstances that always “push” the value in the same direction, such as friction, will be systematic errors.

An experiment will always contain both random and systematic errors. For example, suppose we measurement the time it takes a ball to fall from the roof of Meyer and compare it to the time estimate from the acceleration of gravity. We will start a stopwatch just as we drop it off the roof, and stop the watch when we see it bounce. One source of error is our reaction time. This is a random error: we may delay too long in starting the watch, or delay too long in stopping the watch. In practice, the reaction times will not be the same in repeated measurements; our measured times will be randomly distributed around the true time. Another source of error will be air resistance. This will always cause the time of the ball’s fall to increase. This is a systematic error since it will always add an error in the same direction.

## 6 The Mean and Standard Deviation

The mean, or average, of a set of data is given as

$$\bar{x} = \frac{\sum x_i}{N} \quad (3)$$

where  $N$  is the total number of data points and  $x_i$  are the data, and the numerator is the sum of all the numbers. The standard deviation of a set of numbers is a measure of how much variation there is in the data from the average value. The data set 4, 4.1, 3.9, 3.8, 4.2 will have a smaller standard deviation than 1, 2, 4, 6, 7, even the average,  $\bar{x}$ , is 4 for both. The standard deviation,  $\sigma$  (the greek letter sigma), is

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}} \quad (4)$$

The numerator is a sum of the square root of the deviation from the mean for all data points,  $x_i$ . This is divided by the total number of points,  $N$ , minus one. The standard deviation is rarely calculated by hand. It can be implemented in Excel with the function STDEV(data), in Matlab with std(data), and on free online standard-deviation calculators.

## 7 Propagation of Uncertainties

Often a desired quantity  $q$  is a function of a number of measured quantities  $x, y, \dots, z$ , each with their own uncertainties,  $\delta x, \delta y, \dots, \delta z$ . How the uncertainties in these quantities affect the uncertainty in  $q$  is called error propagation.

### 7.1 Sums and Differences

Consider a quantity such that  $q = \pm x \pm y$ , and the uncertainties in  $x, y$  are known to be  $\delta x, \delta y$ . Then,  $q$  can be as big as

$$\pm x + \delta x \pm y + \delta y \quad (5)$$

or as small as

$$\pm x - \delta x \pm y - \delta y. \quad (6)$$

Or,  $q$  ranges from

$$(\pm x \pm y) + (\delta x + \delta y) \quad (7)$$

to

$$(\pm x \pm y) - (\delta x + \delta y) \quad (8)$$

The formula is a good approximation in many cases, but for a large number of variables or high uncertainties, it is likely to overestimate the uncertainty. This can be defined more carefully in statistics, and in appropriate situations, the uncertainty in  $q$  is obtained by adding the uncertainties in quadrature.

$$\delta q = \sqrt{(\delta x)^2 + (\delta y)^2 + (\delta z)^2} \quad (9)$$

For example, if the length of an object,  $L$ , is the sum of three lengths

$$L = 12 \pm 3\text{cm} + 3 \pm 1\text{cm} + 30 \pm 5\text{cm} \quad (10)$$

Then  $L$  is equal to

$$\begin{aligned} L &= 12 + 3 + 30 \pm \sqrt{3^2 + 1^2 + 5^2} \text{ cm} \\ &= 45 \pm 6 \text{ cm} \end{aligned}$$

### 7.2 Products

If  $q = xy$ , then  $q$  can be as big as

$$(x + \delta x)(y + \delta y) \quad (11)$$

$$= xy + (\delta x)y + (\delta y)x + \delta x\delta y \quad (12)$$

or as small as

$$(x - \delta x)(y - \delta y) \quad (13)$$

$$= xy - (\delta x)y - (\delta y)x + \delta x\delta y \quad (14)$$

Dropping the very small  $\delta x\delta y$  terms, the uncertainty  $\delta q$  is  $y\delta x + x\delta y$ . The quantities  $q, x$ , and  $y$  may not all have the same units, so it is important to use dimensionless quantities. Since  $q = xy$ , we can find the fractional uncertainty in  $q$ ;

$$\frac{\delta q}{q} = \frac{y\delta x + x\delta y}{xy} = \frac{\delta x}{x} + \frac{\delta y}{y} \quad (15)$$

Or, in many statistical situations,

$$\frac{\delta q}{|q|} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2 + \left(\frac{\delta z}{z}\right)^2} \quad (16)$$

For example,  $d = vt$ . If  $v = 20 \pm 3\text{m/s}$  and  $t = 4 \pm 2\text{sec}$ , then

$$\begin{aligned} d &= (20\text{m/s})(4\text{s}) \pm (20\text{m/s})(4\text{s})\sqrt{\left(\frac{3\text{m/s}}{20\text{m/s}}\right)^2 + \left(\frac{2\text{s}}{4\text{s}}\right)^2} \\ &= 80 \pm 42 \text{ m/s} \end{aligned}$$

### 7.3 Quotients

We can treat a quotient  $q = \frac{x}{y}$  much like a product;  $x * \frac{1}{y}$ . The uncertainty can be as big as

$$\frac{x + \delta x}{y - \delta y} \quad (17)$$

or as small as

$$\frac{x - \delta x}{y + \delta y} \quad (18)$$

The same type of analysis can be carried out as in the Products section, leaving us with the same equation:

$$\frac{\delta q}{|q|} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2} \quad (19)$$

### 7.4 Measured Quantity Raised to an exact power

If  $q = x^n$  and  $n$  is known exactly, then

$$\frac{\delta q}{|q|} = |n| \frac{\delta x}{|x|} \quad (20)$$

This equation is a consequence of the binomial theorem, and the derivation requires some mathematical sophistication. A simple way to understand this formula is to remember the rule for products,

$$\frac{\delta q}{q} = \frac{\delta x}{x} + \frac{\delta y}{y} \quad (21)$$

So if we have a formula  $q = xx$ , that is,  $q = x^2$ , then

$$\frac{\delta q}{q} = \frac{\delta x}{x} + \frac{\delta x}{x} = 2 \frac{\delta x}{x} \quad (22)$$

This formula clearly extends for any number  $n$ .

## 8 Linear Regressions

A more general problem is to measure two quantities that are related linearly. That is, the relationship between the two quantities follows  $y = ax + b$ . Least squares fitting can be computed on the computer or calculator. Excel, Matlab, LoggerPro, Mathematica, or a TI-83 or higher calculator can all solve for  $a$  and  $b$  given a data set of at least two points. It is important to understand the error in the linear fit. Often, a program will provide error bounds for the coefficients,  $a$  and  $b$ . These error bounds should be included in the equation. Regression software will also provide the  $r^2$  value for the fit.  $r^2$  is the correlation coefficient. It is a number ranging from 0 to 1, and gives the quality of the least squares fitting to the original data. That is to say, that a correlation coefficient of 1 means that the least squares fit passes exactly through every data point. A correlation coefficient of 0 means the least squares line is entirely uncorrelated to the data. The value of the correlation coefficient will provide some measure to the precision of the data set and least squares fit. When fitting a model  $y = ax + b$ , the quantity  $x$  should be the variable with the smaller uncertainty, and the variable  $y$  should have the larger uncertainty.