

## Analyzing Data: Statistics, Significant Digits and Errors

### Average

The *average* is also known as the *mean*. This is frequently denoted by the variable letter with a bar drawn over it,  $\bar{x}$ . Take care not to confuse this with the symbol for a vector which is similar. The equation for finding the mean of n values is

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

This means that if you have n values for some measurement, add them all together and then divide by n. The Greek letter sigma,  $\sum$ , means "take the sum of". For example, if you make six measurements of the length of an object and you get  $x = 2.45\text{m}, 2.46\text{m}, 2.43\text{m}, 2.44\text{m}, 2.45\text{m},$  and  $2.44\text{m}$ , the average would be

$$\bar{x} = \frac{1}{6}(2.45 + 2.46 + 2.43 + 2.44 + 2.45 + 2.44) \text{ m} = \frac{1}{6}(14.67 \text{ m}) = 2.445 \text{ m}.$$

Note that the sum has four significant digits and so does the average.

There is no way to guarantee the accuracy of any particular measurement. Therefore it is always a good idea to make many measurements and average the results.

### Deviation

The *deviation* is the absolute value of the difference between each measurement and the mean value. There will be as many deviations as there are values of x and the deviations have the same units as x:

$$|d_i| = |\bar{x} - x_i|.$$

The *mean deviation* is just the average of the deviations:

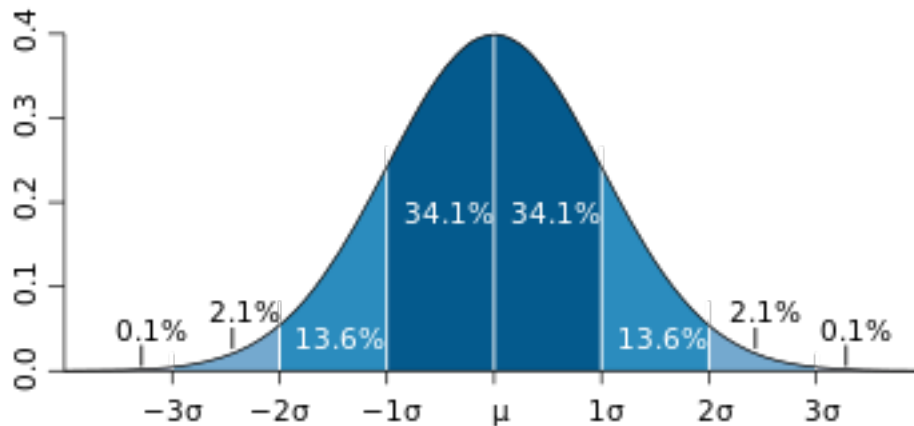
$$|\bar{d}| = \frac{1}{n} \sum_{i=1}^n |\bar{x} - x_i|.$$

The mean deviation gives you a rough idea of how consistent your measurements were. If an experiment is well designed and carefully done, the average value will be correct and the mean deviation will be small. If the experiment is well designed but poorly carried out, the average may still be correct, but the mean deviation will be large.

A more sophisticated statistical measure of the spread of a set of numbers is the *standard deviation*. To calculate it, find the square root of the average of the deviations squared, or root-mean-square deviation. Another practical advantage of this quantity is that it does not require taking absolute values since it uses the square of the deviation. The standard deviation,  $\sigma$ , is given by:

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

68.2 % of your values will fall within  $\pm$  one standard deviation of the mean value. 95% and 99.5% of the values will fall within  $\pm$  two and three standard deviations of the mean respectively.



[http://en.wikipedia.org/wiki/68-95-99.7\\_rule](http://en.wikipedia.org/wiki/68-95-99.7_rule)

### The Slope of a Line

Frequently you will graph one variable as a function of another. The variable that the experimenter controls is called the independent variable and is almost always plotted on the x (horizontal) axis. The values of the variable that is measured, called the dependent variable, are the plotted on the y (vertical) axis. When the instructions say to plot "A versus B", or "A as a function of B", the values for B go on the x axis and the values for A go on the y axis.

Many of the functions that you will plot in this course are linear, that is the data will fall on a straight line. The equation for such a line is  $y = mx + b$ , where  $m$  is the slope and  $b$  is the y-intercept, that is, the value of  $y$  when  $x = 0$ . The slope of this line,  $m$ , tells you how  $y$  changes with respect to  $x$ . It is, in other words, the derivative of  $y$  with respect to  $x$ ,  $\frac{dy}{dx}$ .

The equation for the least squares slope is:

$$m = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{N(\sum x_i^2) - (\sum x_i)^2}$$

The y-intercept is given by:

$$b = \bar{y} - m\bar{x}.$$

The slope units will be the units of y divided by the units of x. The y intercept will have the units of the y values.

Another useful statistic is the uncertainty in the slope or intercept,  $\Delta m$ . The uncertainty is also called the standard error of the estimate or the standard error of the slope. It will have the same units as the slope. If you repeat your experiment you can usually be 95% confident that the result you get for the slope will be within the range of  $m \pm 2\Delta m$ .

### **Significant digits**

The number of digits in a measurement indicates how precise the measurement is. The number of significant digits in a number is not the same as the number of decimal places. For example, both 14.1 and 0.000141 have three significant digits. Zeroes to the left of the number are not significant; zeroes to the right of the number are significant. Therefore 0.00025 has two significant digits and 25.00 has four. The last digit in any measurement is usually uncertain.

There are two different rules to follow when you are doing arithmetic; one is used for addition and subtraction while a separate rule is used for multiplication and division.

### **Addition and Subtraction**

When you add or subtract, the precision of the answer is determined by the least precise measurement.

For example:

$$\begin{array}{r} 11.382 \\ 5.96 \\ + 248.0042 \\ \hline 265.3462 \end{array}$$

5.96 is the least precise measurement therefore the sum should be rounded to 265.35. The answer should be rounded, not truncated.

### **Multiplication and Division**

When you multiply, divide, use trig functions powers or roots, the answer should have as many significant digits as the measurement that had the least.

For example:

$$\sin 16^\circ = 0.28, \quad \text{but} \quad \sin 16.0^\circ = 0.276$$

and

$$(2.2)^2 = 4.8, \quad \text{but} \quad (2.20)^2 = 4.84.$$

$$\begin{array}{r} 3.14159 \\ \times 2.54 \\ \hline 7.979639 \end{array}$$

This final answer should be rounded to three digits, 7.98.

$$\frac{3.22}{14.673} = 0.219$$

The final answer can only have three significant digits.

Exact integers are considered to have an infinite number of significant digits.  
For example:

$$r = \frac{d}{2} = \frac{5.228}{2} = 2.614$$

The final answer will have four significant digits.

When you do an involved calculation with your calculator, carry out all the arithmetic and then round the final answer. If you round your values before you reach the final answer rounding errors may accumulate.

### **Errors**

There are two types of errors that can affect the results of an experiment. *Random errors* affect one measurement at a time. The effect of random error can be reduced by making many measurements of the same variable and averaging the results.

*Systematic errors* are caused by a flaw in the experimental design, the measuring apparatus, or the experimenter's procedure. All measurements are affected so averaging does not improve the accuracy of the final results.

### **Percent error**

Use percent error to compare a result with the accepted value for some quantity. It is the absolute value of the difference between the two values divided by the accepted value and expressed as a percentage.

$$\% \text{ error} = \frac{|\text{accepted value} - \text{experimental value}|}{|\text{accepted value}|} \times 100\%.$$

### **Percent Difference**

Use percent difference to compare two experimental results. It is the absolute value of the difference of the results divided by the average of the results and expressed as a percentage.

$$\% \text{ difference} = \frac{|\text{difference between results}|}{|\text{average result}|} \times 100\%.$$