Uncertainty in Measurement

When scientists collect data, they rarely make just one measurement and leave it at that. Random error and unanticipated events can cause a single data point to be somewhat non-representative. Thus, multiple measurements are usually taken, and the "best value" is represented using a variety of approaches. Mean, median, and mode are measurement of central tendency. The degree to which data tend to be distributed around the mean value is known as variation or dispersion. A variety of measures of dispersion or variation around a mean are available, each with its own peculiar idiosyncrasy. The most common measures are the following: range, mean deviation, standard deviation, semi-interquartile range, and 10-90 percentile range. In this brief article we will examine only those measures most suitable for the purposes of an introductory lab experience.

As a scientist or an engineer you need to know how to deal effectively with uncertainty in measurement. Uncertainty in measurement is not to be confused with significant digits that have their own rules of uncertainty (see *Significant Figures in Measurement and Computation* elsewhere in this *Handbook*).

When one reads that some measurement has the value and absolute error of, say, $d = 3.215m \pm .003m$, how is this to be interpreted? The value $3.215m$ is known as the arithmetic mean (as opposed to weighted mean, geometric mean, harmonic mean, etc.) It is sometimes loosely called the average (but there are different types of averages as well). The arithmetic mean of N measurements is defined as follows:

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

Unless the person who provides the information indicates what the absolute error represents, there is usually no way of knowing. Does the indicated error represent the **range** of the values measured? That is, does this merely indicate that the "true" value of d is somewhere between $3.212m$ and $3.218m$? It could imply this, but other interpretations are also possible.

In another case, $\pm 0.003m$ could represent the **mean deviation**. The mean deviation is the mean of the absolute deviations of a set of data about the data's mean. The mean deviation is defined mathematically as

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

where $\bar{x}$ is the arithmetic mean of the distribution.

In yet another case $\pm 0.003m$ could represent one **standard deviation**. This would then represent a probability that there is a 68% chance ($\pm 1\sigma$ or ± 1standard deviation) that the "best" value is within the range of $3.212m$ and $3.218m$. Such an interpretation is not strictly justified unless more than 30 data points have been used to calculate the mean value and standard deviation. In this interpretation data must also be "normally distributed" around the mean. That is, the mean, median, and mode of the data must be essentially the same. There must be a "bell shaped" distribution of data that are symmetrically distributed around the mean. Most data collected in introductory labs fail to meet these criteria.

Whatever form of error representation you use in lab, you must be consistent in your expressions of uncertainty in measurement. You must also declare what your error represents. For the purposes of these introductory labs, the following forms of uncertainty in measurement will be used if called for.

Example 1: Range – You use a meter stick with divisions marked off in $1mm$ units. The measurements of length appear to be somewhere between $93mm$ and $94mm$. You would best represent the value of the length with uncertainty as follows stating without any ambiguity that the true value of the length measured lay somewhere between $93mm$ and $94mm$.

$$93.5mm \pm 0.5mm$$

Example 2: Mean Deviation – You make six observations using an electronic photogate timer to measure nearly identical falls of a ball with the following results:

Observation 1: 6.556s
Observation 2: 6.576s
Observation 3: 6.630s
Observation 4: 6.587s
Observation 5: 6.575s
Observation 6: 6.602s

The value of the mean deviation is calculated thus:

$$\bar{x} = 6.588s$$
$$|x_1 - \bar{x}| = 0.022s$$
$$|x_2 - \bar{x}| = 0.012s$$
$$|x_3 - \bar{x}| = 0.042s$$
$$|x_4 - \bar{x}| = 0.001s$$
$$|x_5 - \bar{x}| = 0.013s$$
$$|x_6 - \bar{x}| = 0.014s$$

Sum of deviations = 0.104s
Mean deviation = 0.017s

The mean of the data along with mean deviation are accurately stated as $6.588s \pm 0.017s$

Example 3: Standard Deviation – Let's say you have made 30 measurements using your electronic timer. For the purposes of these introductory labs we will use the mean ±1standard deviation (or ±1σ) to represent the error. This process makes three major assumptions – that the variation in the data comes only from random error, that 30 or so data points are sufficient to calculate a good standard deviation, and that data constitute a "normal" (bell-shaped distribution) around the mean. None of these assumptions might be correct in a particular situation, so it might be a good idea to examine the distribution of data to make certain that data have the standard symmetrical "bell shaped" curve. Thus, interpretation of the standard deviation must be done with great caution.

Standard deviation of N data points can be readily calculated with the use of the following formula:

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

Most scientific calculators will, given the data points, readily calculate the standard deviation with ease. Lab students should use this form to express uncertainty in repeated measures throughout these introductory lab activities when called for. Due caution should be observed in the use of significant digits.