4. Uncertaintiy Analysis

A. How can we describe our process or products when they are random?

Now we can make one more definition for error, Error in a measurement means the inevitable uncertainty that attends all measurements. This definition emphasizes that the error is a reality in physical word. The amount of the error in engineering disciplines are defined with a concept of engineering tolerance. The engineering tolerance (note that it is different form mechanical tolerance) is **the permissible limit or limits** of variation in a physical dimension. Therefore, the representation uncertainty of the experimental data should have at least two components.

- An estimate of expected value, that is mean of a sample
- A tolerance band, this tolerance band is mostly +/-three times of standard deviation because a band of +/- three standard deviation around mean represents 99% of measurements in a sample and it is the engineering confidence level.

Suppose we are drilling a hole on a workpiece with a diameter of 12mm and our contracter's demand is to have pieces with a hole of 12 +/- 0.3 mm. So your sample should have, at worst, mean of 12mm and standart deviation of 0.3 mm.

B. Uncertainty of a measurement

Uncertainty means having a limited knowledge where it is impossible to exactly describe the existing state, a future outcome, or more than one possible outcome. Therefore, a confidence level is critical for the validity of the expression. Refering to definition on engineering tolerance, the uncertainty (u_i) of a dimension (or a variable) x_i is given as

$$u_i = \pm 3\sigma_i$$

This is general case and it can also be used to express the precission of an instrument or a machine.

C. Calculating uncertainty of a population from samples

It is rare that anyone can measure something for an entire population instead, a sample (or several samples) of the population is measured, and the population statistics are estimated from the sample.
The Central Limit Theorem (CLT) is a useful tool when dealing with multiple samples.

Consider a population of random variable *x* (we assume that variations in *x* are purely random – in other words, if we would plot a PDF of variable *x*, it would look Gaussian or normal). The population mean and the population standard deviation are not known, but are instead estimated by taking several samples.

Data	Mean	Data	Mean		Data	Mean
X_1		X_1			X_1	
X_2		X_2			X_2	
X_3		X_3			X_3	
X_4	$\bar{X}_I$	X_4	$\bar{X}_{II}$	....	X_4	$\bar{X}_N$
X_5		X_5			X_5	
X_6		X_6			X_6	
X_7		X_7			X_7	
X_8		X_8			X_8	

The sample mean for sample I is defined as

Topic 3.Uncertainty

$$\bar{X}_I = \frac{\sum_{i=1}^{n} x_i}{n}$$

where index I = 1, 2, 3,..,N. In other words, we calculate a sample mean in the usual fashion – an average value for each sample 1 through N. We collect all *N* values of sample mean and treat this collection as a sample itself (we call it the sample of the sample means). The ***sample of the sample means*** consists of *N* data points: 1 2 3 , , , ... *N*. We perform standard statistical analyses on this sample of the sample means:

$$\left(\bar{X}_I\right)_{mean} = \frac{\sum_{I=1}^{N} \bar{X}_I}{N}$$

The mean of the sample means

It should be obvious that the sample mean of the sample means is identically equal to the overall sample mean of all the data points combined (all N samples of n data points each).
Our best estimate of the population mean is thus

$$\mu = \left(\bar{X}_I\right)_{mean}$$

The sample standard deviation of the sample means (also called the standard error of the mean) is calculated by the usual definition of sample standard deviation, but applied to the N sample means:

$$S_{\bar{X}} = \sqrt{\frac{\sum_{I=1}^{N}\left(\bar{X}_I - (\bar{X}_I)_{mean}\right)^2}{N-1}}$$

The Central Limit Theorem (CLT) is stated as follows: As N approaches infinity, the sample Standard deviation of the sample means approaches the overall sample standard deviation divided by the square root of N:

$$\sigma = \frac{S_{\bar{X}}}{\sqrt{N}}$$

However, as n gets large, the sample standard deviation approaches the population standard deviation. As *n* and *N* get large, the PDF of the sample of the sample means is Gaussian even if the underlying population is *not* Gaussian.

C. Calculating combined uncertainties

Experimental uncertainty analysis provides a method for predicting the uncertainty of a calculated variable based on its component uncertainties. This analysis is also called the propagation of uncertainty

It is most common that the variables to be monitored in processes are not measurable but the variables to be monitored can be selected so that hety can be calculated form measurable ones. In scuh cases propogation of uncertainty is exploited.

we measure N physical quantities (or variables, like voltage, resistance, power, torque etc.), x_1, x_2, ….,x_N. Suppose that each of these quantities has a known experimental uncertainty associated with it: $x_i = \bar{x}_i \pm u_{x_i}$

Furthermore, unless otherwise specified, each of these uncertainties (u_1, u_2, ….,u_N) has a confidence level of 95%. Since the variables are components of the calculated quantity, we call the uncertainties **component uncertainties**.

Some new variable, R, is a function of these measured quantities
$R(x_1, x_1, ….,x_N)$.
The goal in experimental uncertainty analysis is to estimate the uncertainty in R to the same confidence level as that of the component uncertainties, i.e., we want to report R as

Topic 3.Uncertainty

$R = \overline{R}_i \pm u_R$, where u is the predicted uncertainty on variable R.

There are two types of uncertainty on variable R:

Maximum uncertainty – We define the maximum uncertainty on variable R as $u_{R,max}$

$$u_{R,\max} = \sum_{i=1}^{i=N} \left| u_{x_i} \frac{\partial R}{\partial x_i} \right|$$

Because of the absolute value signs, this expression assumes that all the errors in the component variable x_i measurements are such that the error in R is always the same sign. Such a case would be highly unlikely, especially for a large number of variables (large N), because some of the errors would be positive and some negative, and the errors would cancel each other out somewhat. In other words, this is a worst case scenario.

The root of the sum of the squares uncertainty is more realistic than the maximum uncertainty since it is unlikely that the maximum error will occur on all component variables simultaneously.

$$u_{R,RSS} = \sqrt{\sum_{i=1}^{i=N} \left(u_{x_i} \frac{\partial R}{\partial x_i} \right)^2}$$

RSS uncertainty is the engineering standard, and the usual notation is to set u_R is equal to the u_{RSS}.

The root-of-the-sum-of-the-squares (RSS) concept is also useful when one needs to combine elemental uncertainties defined as precision uncertainties, bias uncertainties, calibration uncertainties, etc. Consider K elemental uncertainties:
To obtain an overall estimate of the uncertainty in the measurement of x, the standard convention is to use

$$u_x = \sqrt{\sum_{i=1}^{i=K} u_i^2}$$

Like

$$u_x = \sqrt{\left(u_{\text{systematic}}\right)^2 + \left(u_{\text{random}}\right)^2}$$

C. Expressing uncertainties

Experimental uncertainties should be rounded to one significant figure. Experimental uncertainties are, by nature, inexact. Uncertainties are almost always quoted to one significant digit (example: ±0.05 s). If the uncertainty starts with a one, some scientists quote the uncertainty to two significant digits (example: ±0.0012 kg).
Wrong: 52.3 cm ± 4.1 cm

Correct: 52 cm ± 4 cm
Always round the experimental measurement or result to the same decimal place as the uncertainty. It would be confusing (and perhaps dishonest) to suggest that you knew the digit in the hundredths (or thousandths) place when you admit that you unsure of the tenths place.
Wrong: 1.237 s ± 0.1 s

Correct: 1.2 s ± 0.1 s