

Metrology Musings: Measurement Uncertainty for Anyone

By [Bob Lutz](#), AASHTO re:source Manager
 Posted: May 2012

"Do not imagine that mathematics is hard and crabbed, and repulsive to common sense. It is merely the etherealisation of common sense."
Lord Kelvin, 1824-1907

In my last article ("[Uncertainty? There's An App For That](#)") I promised that we would further explore the numbers and the math that goes into estimating measurement uncertainty. As Churchill Eisenhart once said, "A quantitative result without any kind of uncertainty estimate is not only useless, it is dangerous because it can be misused." Unlike bias, or systematic error, measurement uncertainty is always there and can never be eliminated. You have to know what it is and then deal with it. If you calibrate any of your own measurement equipment then you should be familiar with the process of estimating uncertainty. Even if you don't calibrate any equipment, it helps to understand the development of measurement uncertainty estimates so you can evaluate calibration certificates from your service providers.

Uncertainty of measurement also determines whether a measurement process meets the needs of the application. We take measurements in our laboratories almost every day. We use these measurement results for quality control, quality assurance, judging engineering designs, and approval of materials for use. The measurements are important...but no measurement is perfect. Therefore, uncertainty of measurement characterizes the doubt (or confidence, depending on your perspective) that exists about the measurement so that we can conclude something about its quality.

Now, I won't lie and tell you that estimating measurement uncertainty is as easy as the "ciphering" Jethro Bodine (any *Beverly Hillbilly* fans?) used to do as he proudly displayed his "gozintas" skills - "three gozinta three one time, three gozinta six two times," and so on. No, it's not that simple, but don't be intimidated – I am going to walk you through the real calibration process, and resulting uncertainty budget, for a stopwatch used by AASHTO re:source.

The Uncertainty Budget As A Tool

Measurement uncertainty budgets are used to identify and quantify the sources of uncertainty in our measurements. As I discuss the calibration of an AASHTO re:source stopwatch, I will explain how the **standard uncertainty** for each source of uncertainty was estimated, how the **combined uncertainty** was calculated, how the **expanded uncertainty** was calculated, and finally what we did with all of the information generated from this uncertainty budget.

Types of Data Distributions

To understand how to calculate the standard uncertainty for each source of measurement uncertainty, you first need to understand a little bit about probability distributions, because the spread (distribution) of a set of data can take different shapes. Let's illustrate with some simple examples.

If you roll two dice enough times, the resulting rolls will follow the normal distribution pattern shown in Figure 1. (There are 36 possible combinations.) Values are more likely to fall near the average than farther away.

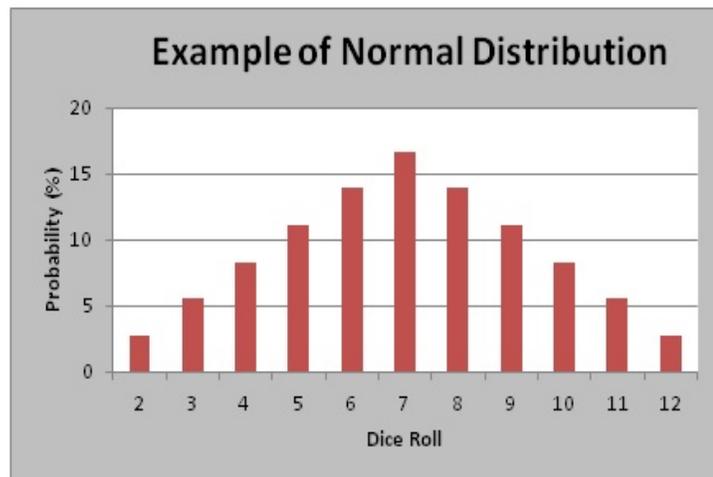


Figure 1: Dice Roll Probability (Two Dice)

You have a higher probability of rolling a seven (16.67%) than any other number because there are more ways to roll a seven than any other number: 1-6, 2-5, 3-4, 4-3, 5-2, and 6-1. ($6 \text{ combinations} / 36 \text{ possible combinations} = 16.67\%$) There is only one way to roll a two, 1-1, and only one way to roll a twelve, 6-6. ($1 \text{ combination} / 36 \text{ possible combinations} = 2.78\%$) See, it's not really "lucky" seven after all – it's really "highest probability" seven.

Note: Many distributions fall on a normal curve, especially when large samples of data are considered. These normal distributions include height, weight, IQ, and other standardized test scores, like the SATs.

For a normal distribution, we know (see Figure 2):

- 68% of the population falls within one standard deviation (± 1 standard deviation) from the average value.
- 95% of the population falls within two standard deviations (± 2 standard deviations) from the average value.
- 99% of the population falls within three standard deviations (± 3 standard deviations) from the average value.

Applying this knowledge to our dice data, we can say that:

- The average roll is a 7;
- The standard deviation is 2;
- There is a 68% probability that any roll of the dice will be between 5 and 9 [7 ± 2 (1 standard deviation)];
- There is a 95% probability that any roll of the dice will be between 3 and 11 [7 ± 4 (2 standard deviations)].

We will talk more about these probabilities later, but keep this question in mind: If you're betting on the *next* roll, do you want a 68% chance of being correct, or a 95% chance?

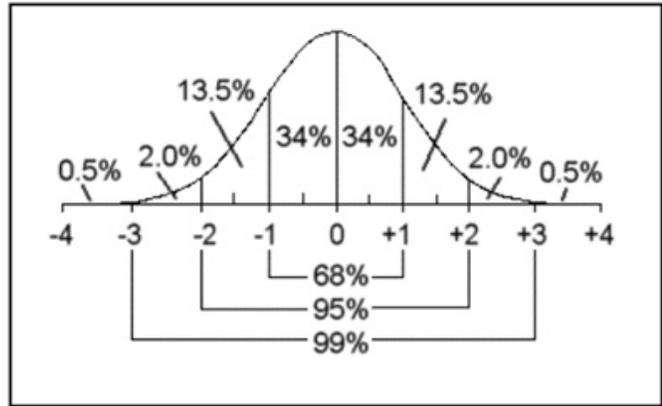


Figure 2: Normal Distribution

Now, take away one of the two dice and start rolling just one die – what will that distribution look like? Will it change? Yes, it will. Now each number from 1 – 6 has the same chance, the same probability, of being rolled ($1/6 = 16.67\%$). If you roll one die enough times, the results will look like the pattern in Figure 3.

This is called a uniform, or rectangular, distribution, where all outcomes have an equal probability of occurring. Calculating one standard deviation for this type of distribution is a little tricky...stay tuned. On to our calibration.

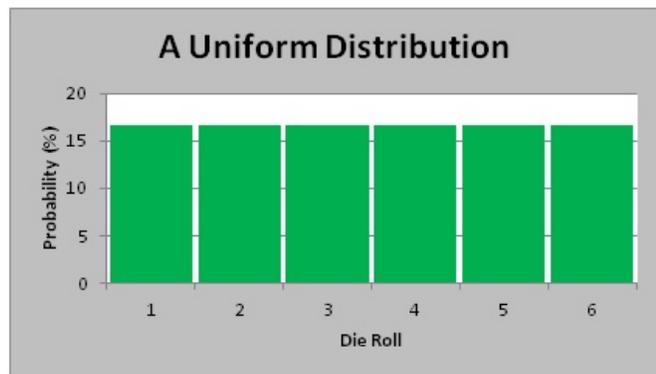


Figure 3: Uniform Distribution

The Hillbilly Experiment

Step 1: Organizing the Experiment

We decided to keep this simple: calibrate one stopwatch using a direct comparison, which means we would compare our stopwatch readings to a traceable audio time signal produced by the National Institute of Standards and Technology (NIST).

Note: NIST continually broadcasts the time using shortwave radio signals from two radio stations: WWV near Fort Collins, Colorado, and WWVH in Kauai, Hawaii. For convenience, they also maintain two telephone numbers that allow you to listen to the time signal: 303-499-7111 (Colorado) and 808-335-4363 (Hawaii). Refer to "[Standardizing Timers and Stopwatches](#)" for the reason why the time displayed by NIST on the internet (<http://www.time.gov>) is not a suitable reference standard.

We selected four testers to perform the comparison ten times by:

1. Calling the phone number (303-499-7111) from a land line;
2. Listening to the signal and waiting for the mark (every minute the voice says, "At the tone, x hours, y minutes, coordinated universal time," followed by a beep);
3. Pressing the start button at the sound of the beep;

4. Hanging up the phone (the length of the phone call is limited to about two minutes) and allowing the stopwatch to run;
5. Repeating steps 1 – 3 for an elapsed interval of 5 minutes (300 seconds).

Step 2: Collecting the Data

Table 1 summarizes the data from our stopwatch calibration experiment. You can see that the average values were slightly different, and the standard deviations for each tester varied quite a bit. None of this should be surprising, however; we expected that the reaction times to the audio signal would vary for each of our human testers. That will lead us to the first source of measurement uncertainty in our budget.

	Tester 1	Tester 2	Tester 3	Tester 4
Reading 1	300.07	299.94	300.07	300.08
Reading 2	300.13	299.9	300.07	300.15
Reading 3	299.96	300.07	300.03	300.27
Reading 4	300.10	299.81	300.06	299.94
Reading 5	300.07	300.06	300.00	300.15
Reading 6	300.03	300.02	300.03	300.16
Reading 7	299.97	299.88	300.08	299.89
Reading 8	300.02	300.09	300.06	300.09
Reading 9	300.00	299.95	299.95	300.21
Reading 10	299.95	300.01	300.06	299.89
Average	300.030	299.973	300.04	300.08
Std. Deviation	0.058	0.092	0.040	0.134

Table 1: Stopwatch Calibration Data

Step 3: Constructing the Measurement Uncertainty Budget

a. Uncertainty due to human reaction time

We pooled the data into one big set of forty measurements and graphed the distribution of data (Figure 4). As expected, the result was a normal distribution.

We plugged the raw data into Excel® and calculated one standard deviation – in this case it was 0.094 s. Now we can begin to construct our uncertainty budget:

Source of uncertainty	Standard uncertainty (s)
Human reaction time	0.094

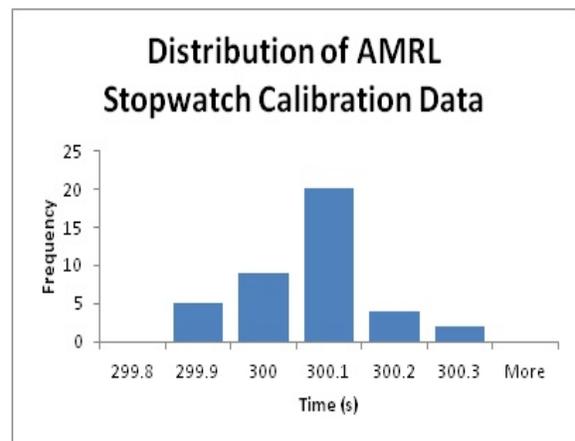


Figure 4: Distribution of Stopwatch Data

b. Uncertainty of the measurement reference standard

NIST has stated that callers in the continental United States using ordinary land lines can expect signal delays (the time from when the signal was produced until it reached your ear) of less than 30 ms (0.030 s) when dialing (303) 499-7111, and that these delays should be very repeatable from phone call to phone call.

Note: NIST has also stated that calls made over wireless or VOIP networks can have signal transmission delays as high as 150 ms (0.150 s). Calls made from outside the continental United States might be routed through a communications satellite, which could result in delays of 250 ms (0.250 s).

Say what?- A second is defined as the time interval required for 9,192,631,770 transitions between two energy states of the cesium atom to take place. The atomic definition of the second, together with current technology, allows it to be measured with much smaller uncertainties than any other SI unit. NIST can measure a second with an uncertainty of less than 1 part in 1015, or more than 1 billion times smaller than the uncertainties required for the calibration in our example.

Standard Uncertainty – What Is It? The uncertainty for each source should be expressed at the same level, so that we can standardize each component and then perform our math. Think of standard uncertainty as plus / minus one standard deviation.

In this case we don't have all the raw data that led to the estimate of 0.030 s, and we don't know anything about the distribution of the data, so we have to assume that all probabilities are equal and that it's a uniform distribution. So how do we calculate the standard uncertainty for a uniform distribution??? Excel® can't do it. Well, if $\pm a$ is the uncertainty of the reference standard (± 0.030 s in our example), the variance due to this uncertainty is equal to: $a^2/3$, and **its standard deviation is equal to: $a/\sqrt{3}$.**



$$0.030 \text{ s} / \sqrt{3} = 0.017 \text{ s}$$

We add that to our uncertainty budget:

Source of uncertainty	Standard uncertainty (s)
Human reaction time	0.094
Reference standard (telephone delay)	0.017

c. Uncertainty due to resolution of device under test

Resolution refers to the smallest interval the stopwatch can display. In this example, the stopwatch displays two digits to the right of the decimal, so its resolution is 0.01 s (1/100 of a second). Makers of stopwatches have to set realistic limits and expectations for the resolution, so they generally stop at two decimal places. That means our stopwatch can display a value of 900.17 s, or 900.18 s, but cannot display a value of 900.174 s. A value of 900.17 could have been as low as 900.165 and as high 900.175, so the limit to the resolution is ± 0.005 s. Again, the distribution of the data is rectangular, so the standard uncertainty for this component is:

$$0.005 \text{ s} / \sqrt{3} = 0.003 \text{ s}$$

Now we add that to our uncertainty budget:

Source of uncertainty	Standard uncertainty (s)
Human reaction time	0.094
Reference standard (telephone delay)	0.017
Resolution of stopwatch	0.003

Step 4: Combining the standard uncertainties

This step is so easy Jethro could perform it. The combined standard uncertainty is calculated by a method called “root sum of the squares.” This means “square each standard uncertainty, add them together, and then calculate the square root of the result.”

Source of uncertainty	Standard uncertainty (s)	Standard uncertainty squared
Human reaction time	0.094	0.008836
Reference standard (telephone delay)	0.017	0.000289
Resolution of stopwatch	0.003	0.000009
Sum of squares		0.009134
Combined uncertainty = Square root sum of squares		0.096 s

The combined measurement uncertainty is 0.096 seconds. You can see that the result is dominated by human reaction times and that the telephone signal delay and resolution of the stopwatch were not significant. If you wanted to improve this process you should focus on that source of uncertainty.

Step 5: Calculating expanded uncertainty

Remember our betting question on the next roll of the dice? Surely you'd like odds of 95% over odds of 68%. What kind of odds do you want in your laboratory measurements?

So we take our newly-calibrated stopwatch to the lab and measure a time interval for an important test. The stopwatch displays 75.28 s. We realize that measurements cannot ever be perfect, and that the absolute true time interval cannot ever be known...but it can be estimated.

The combined standard uncertainty from our experiment above is ± 0.096 s. (We will round this to ± 0.10 s). Using the same distribution and probability numbers as we saw above (and repeated in Figure 5), we can say that a very good estimate of "truth" is 75.28 ± 0.10 s. That means there is a 68% chance that "truth" falls within this range. But 68% probability doesn't allow us to sleep well at night, especially if our time measurement is a critical part of the test. What should we do?

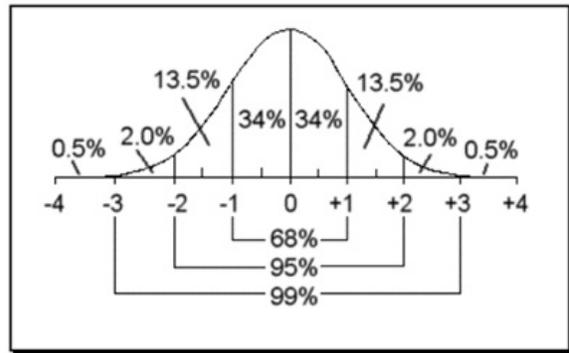


Figure 5: Normal Distribution

Most people feel pretty comfortable with 95% probability, so we simply multiply our uncertainty estimate by two. That multiplier, designated as k , is referred to as the "coverage factor."

We could multiply by three ($k=3$) but that only gets us an additional 4% comfort level, so $k=2$ is the standard number used. Now we can say with great certainty (95%) that the "truth" is 75.28 ± 0.20 s.

Source of uncertainty	Standard uncertainty (s)
Human reaction time	0.094
Telephone delay deviation	0.017
Resolution of stopwatch	0.003
Combined uncertainty	
	0.10 s
Expanded uncertainty, $k= 2$, 95% confidence level	0.20 s

Step 6 – The Last Step!

We don't go to all this trouble just to come up with an estimate for expanded uncertainty and then file it away. We first use this estimate to evaluate fitness for purpose by comparing the expanded measurement uncertainty to the required test tolerances. In the example above, can you live with uncertainty of measurement of nearly 0.2 s? If the time of asphalt flow must be measured to the nearest 0.1 s, could you measure it with confidence? This stopwatch may be acceptable for other tests but it may not be "fit for purpose" in this particular application. The information generated from your uncertainty budget, or obtained from your calibration certificate, can help you decide. AASHTO resource uses the information generated from our uncertainty budgets to make informed decisions when we check our customers' testing and measuring equipment.

If you find that the expanded uncertainty is greater than you'd like, or need, then you can use the uncertainty budget to identify the sources that contribute the most, and then figure out ways to reduce them. And now that you understand the basics of uncertainty budgets, you might want to request them from your calibration providers and examine them. Tell them Jethro wants to know.