Table of Contents

I. Introduction

"Testing" is a term that covers a huge range of activities. Not every test is a measurement. However, for those tests that are measurements and those that include measurements, uncertainty of measurement is an important topic. The following is a simple introduction to the estimation of measurement uncertainty that is applicable to testing in general[1].

The purpose of measurement is to determine the value for a quantity of interest. Examples include the concentration of alcohol in a blood sample, boiling point of water at 1 atmosphere of pressure, the Rockwell hardness of a metal specimen, the tensile strength of a plastic compound, and the length of a metal specimen at 20°C. Calibrations are tests that compare indicated values to input quantities. A calibration is a measurement. Looking something up in a reference book is not measurement. Nominal quantities such as hot, cold, or pretty are not measurements. Measurements are processes that determine quantity values.

Before a measurement can be made, we have to know what we are to measure (the "measurand"), the method and procedure to be used, the test conditions, the measurement devices and systems to be used, and other relevant factors. (See VIM 2.1, GUM 3.1) One of those relevant factors is the measurement uncertainty required. For example, lumber for a dog house does not have to be measured as accurately as piston rods for automobile engines. When we report the results of a measurement, it is important that we report the value and the uncertainty so that they are understandable and relevant to the user.

Every measurement has uncertainty associated with it. Measurement devices, calibration standards, reagents, and tools are not perfect. Environmental conditions, processes, procedures, and people are also imperfect and variable. In order for two measurements to be compared, both must trace back to a common reference. In order for two measurement uncertainty statements to be compared, they must also both trace back to a common reference. The appropriate method of measurement uncertainty calculation depends upon the nature of the test and may be as simple or complicated as necessary to meet requirements. Measurement uncertainty is important not only for calibrations but in any test that involves measurements.

This guide is an introduction to test measurement uncertainty using the method of estimation described in the JCGM 100:2008 *Guide to the Expression of Uncertainty in Measurement* (GUM)[2]. A mechanical testing example is used for illustration in this guide but the methods presented are applicable to many test situations.

II. Definitions[3]

Measurement Uncertainty[4]: a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.

Think of it as a parameter associated with the result of a measurement that characterizes the dispersion of the quantity values that is attributed to the measurand.

Uncertainty in a measurement quantity is a result both of our incomplete knowledge of the value of the measured quantity and of the factors influencing it. There are many possible sources of uncertainty in measurement including[5]:

1) incomplete definition of the quantity being measured;

---

[1] The application of control charts for estimating measurement uncertainty is covered in Appendix A.
[2] The type of analysis required for actual test measurement uncertainty estimation depends upon the nature of the test. (See A2LA R205).
[3] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms", JCM200:2008, BIPM
[4] VIM 2.26
[5] GUM 3.3.2.

2) imperfect realization of the definition of the quantity being measured;

3) non-representative sampling;

4) inadequate knowledge of the effects of environmental conditions on the measurement or imperfect measurement of environmental conditions;

5) personal bias in reading analog instruments, including the effects of parallax;

6) finite resolution or discrimination threshold;

7) inexact values of measurement standards and reference materials;

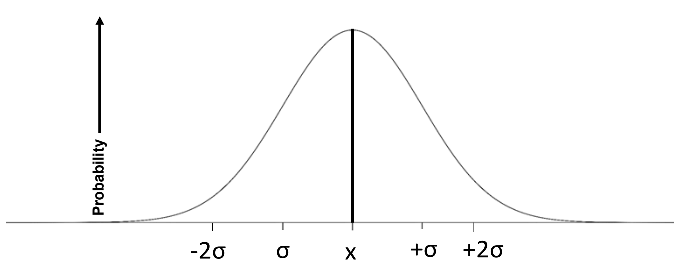
8) inexact values of constants and other parameters obtained from external sources and used in the data-reduction algorithm;

9) approximations and assumptions incorporated in the measurement method and procedure;

10) variations in repeated observations of the measurand under apparently identical conditions.

These sources of uncertainty are not necessarily independent and some or all can contribute to the variations in repeated observations. Not only can uncertainties be introduced by measurement equipment and test methods, but also by the person performing the test, data analysis, the environment, and a host of other factors.

Measurement Uncertainty in Testing: Tests are performed in accordance with test procedures. The use of recognized standard procedures (e.g., ASTM D638) eliminates many potential sources of measurement uncertainty. Definitions, calculations, and other information necessary to evaluate the test data are contained in such test procedures. The procedure addresses test measurement statistics and uncertainty at the level necessary to meet test requirements. Some of the most commonly used terms and concepts follow.



Xi, the expected value. It is the mean of an interval that contains the measured value. There is a 95% probability that the value is between ±2 standard deviations.

Figure 1

Standard Deviation Repeated measurements from a controlled process are described by the Normal (or "Standard") probability distribution that yields an average and standard deviation for the set. The average value is usually taken as the best estimate of the measured quantity. This average is obtained from a number, *n,* of test results according to the formula below.
If we designate each of the test results by the symbol x_i, the following equation gives the **average**, $\bar{x}$, of n test results:

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

The standard deviation, s, characterizes the variability, or spread, in the observed values x_i. It is given by

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \qquad 2)$$

Precision: is closeness of agreement between indications or measured quantity values obtained by repeated measurements on the same or similar objects under specified conditions. Measurement precision is usually expressed as standard deviation, variance, or some other measure of imprecision. Measurement precision is used to define repeatability, reproducibility, and other statistics.

To evaluate test measurement precision, many test procedures require the testing organization to perform a number of repeat measurements and compare the repeatability standard deviation or some other statistic to values specified by the test method.

Control charts (Appendix A) provide important tools for controlling test process precision and bias.

Repeatability[6]: is a condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time.

This involves precision under repeatability conditions, i.e. conditions where test results are obtained with the same method on the same or similar test items in the same laboratory by the same operator using the same equipment at the same location within short intervals of time. Repeatability may be expressed in terms of multiples of the standard deviation.

Repeatability standard deviation is the standard deviation of test results obtained under repeatability conditions.

Reproducibility[7]: is precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on the same or similar test items in different laboratories, by different operators, using different equipment, in different locations, or on different days. Reproducibility may be expressed in terms of multiples of the standard deviation.

Reproducibility standard deviation is the standard deviation of test results obtained under reproducibility conditions. The conditions under which reproducibility is determined should be clearly specified.

Bias[8]: the estimate of a systematic measurement error.

Think of it as the difference between the test results and an accepted reference value. Known biases can be corrected. Bias is often called "systematic error", but corrected biases are not errors. There may be one or more error components, known or unknown, contributing to the bias.

Many test procedures require laboratories to demonstrate that their measurement bias is within prescribed limits by one of the following methods:

1) Reference material: Using an appropriate reference standard or material, the laboratory should perform replicate measurements to form an estimate of its bias, which is the difference between the mean of its test results and the certified value of the standard or material. If the absolute value of this bias is less than twice the reproducibility standard deviation given in the precision statement in the test method, then the laboratory may consider that its bias is under control.

2) Interlaboratory comparison: Laboratories participating in proficiency testing schemes will have available to them data from a large number of laboratories which they can use to estimate the bias of their measurement

[6] VIM 2.20
[7] VIM 2.2.5
[8] VIM 2.18

results. Comparison of the lab mean to the grand mean or other assigned value in such programs, for example, will allow them to demonstrate that their bias is under adequate control.

For many test procedures, bias has not or cannot be evaluated due to the lack of appropriate reference material. In such cases, this fact should be clearly documented.

Trueness (Measurement Accuracy[9]): the closeness of agreement between a measured quantity value and a true quantity value of a measurand.

Think of it as the closeness of agreement between the average value obtained from a large set of test results and an accepted reference value. The measure of trueness is normally expressed in terms of bias.

Measurement Method[10]: a generic description of a logical organization of operations used in a measurement.

It refers to the general description of the measurement such as comparison, substitution, etc.

Measurement Procedure[11]: a detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result.

Think of it as a detailed description of the measurement process. If the laboratory is applying a standard, validated test method, the test method may include definitions for statistical quantities and estimates of precision and bias obtained by interlaboratory comparison during the course of method validation. For example, ASTM test methods are often accompanied by tables of values determined from round-robin reproducibility tests including multiple laboratories and data from repeatability studies performed in a single laboratory.

Measurand: refers to the particular quantity to be measured in a test. Any uncertainty analysis must begin with a clear understanding of the quantity to be measured, the measurand.

Systematic Errors[12]: a component of measurement error that in replicate measurements remains constant or varies in a predictable manner.

Think of these as biases that cause a measurement result to differ from the true value. Taking repeated measurements and averaging them does not improve systematic error. Known systematic errors can be corrected.

Random Errors[13]: a component of measurement error that in replicate measurements varies in an unpredictable manner.

These result from variations in the values of repeated measurements. Taking more repeated measurements generally reduces the random error.

Corrected Value: is the measurement result after systematic effects (biases) are removed.

Intermediate Precision Condition of Measurement[14]: a condition of measurement, out of a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time but may include other conditions involving changes.

---

[9] VIM 2.13
[10] VIM 2.5
[11] VIM 2.6
[12] VIM 2.17
[13] VIM 2.19
[14] VIM 2.23

III. Tensile Strength Example

The ASTM D638 tensile strength at break test can be used to illustrate the application of test measurement uncertainty principles. This mechanical test will also be used to illustrate the GUM method[15].

Tensile strength at break, "S", is defined[16] as the force (called "load" in the ASTM document), F, divided by the cross-sectional area, A, of the test specimen. The design and dimensions of the test specimen, accuracy of the force indicator, and other factors related to the test are clearly identified in the standard and are required to comply with specified tolerances and to be performed in a specified manner. The standard also gives rules and tolerances for other important aspects of the test, including specimen preparation, mounting, conditioning, and speed of testing.

$$S = F/A$$

The cross-sectional area, A, is defined as the thickness, T, of the specimen multiplied times the width, W, when these measurements are made as specified in the standard.

$$A=TW$$

The tensile testing machine stretches the test specimen, continuously measuring the load until the specimen breaks. According to the method, dimensional measurements must be made with an uncertainty of ±0.001 in. or less.

The table below summarizes ASTM D638 test specifications appropriate to this example. The standard addresses a wide range of tests and specimens not addressed here.

Test Specifications		
Load (Force) Indicator	**Thickness**	**Width**
±1% of indicated value	0.13 ±0.02 in.	0.50 ±0.02 in.

Table 1

Let us assume that we perform this test in accordance with the standard. Five specimens are measured for thickness and width, and then tested in the machine. We record the dimensional measurements and the load at break for each (see table below). Then, in accordance with D638, we calculate tensile strength, S, for the test by dividing the maximum indicated load by the average specimen cross-sectional area. S and standard deviation, Sr, are also calculated. Sr is the "in-laboratory standard deviation".

Test Specimen	**Measured Thickness, T**	**Measured Width, W**	**Area, TW**	**Measured Load, F**	**Calculated S**
1	0.124 in	0.499 in	0.0619	830 lb	13409 PSI
2	0.126 in	0.501 in	0.0631	**900 lb (Maximum)**	14263 PSI
3	0.125 in	0.500 in	0.0625	810 lb	12960 PSI
4	0.126 in	0.500 in	0.063	870 lb	13810 PSI
5	0.124 in	0.499 in	0.0619	850 lb	13732 PSI

[15] ASTM D638 does not require the test performer to calculate test measurement uncertainty.
[16] For clarity and consistency in this document, the symbols used in ASTM D638 are not used here. See ASTM D638 A2.24.

Average:	0.125 in	0.500 in	**0.0625**	852.0 lb	**13635 PSI**
	Std. Deviation, S_r:				**485 PSI**
ASTM D638 Tensile Strength = 900/0.0625 = 14,400 PSI					
Table 2.					

The standard method defines repeatability as I_r=2.83S_r, and reproducibility[17] is defined as I_R=2.83S_R in the standard[18]. We calculated Sr for the measurements we made. Tables of S_r and S_R based on round robin studies are provided in the ASTM document for reference.

The standard states that judgments made in accordance with these definitions of repeatability and reproducibility will have an approximate 95% (0.95) probability of being correct. The standard also explains that bias has not been established for this test method because no recognized standards exist.

As illustrated by this example, even though measurement uncertainty may not be calculated for a particular test, standard metrological and statistical definitions and methods apply and that even though systematic errors (bias) remain indeterminate, reliable and consistent test results can be produced and compared among performing organizations.

The same definitions and methods are used in the GUM, along with a few more that were developed specifically for its purposes.

IV. Introduction to the GUM method

The GUM method is not magic. Its application will not produce accurate estimates of measurement uncertainty from bad tests or poor research. What the GUM does provide is a consistent method for estimating measurement uncertainties. These words from that document summarize the situation well:

> Although this Guide provides a framework for assessing uncertainty, it cannot substitute for critical thinking, intellectual honesty and professional skill. The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value. (GUM 3.4.8)

The GUM method is an eight step process:

1. Describe the measured value in terms of your measurement process. (Model the measurement.)
2. List the input quantities
3. Determine the uncertainty for each input quantity
4. Evaluate any covariances/correlations in input quantities
5. Calculate the measured value to report
6. Correctly combine the uncertainty components
7. Multiply the combined uncertainty by a coverage factor
8. Report the result in the proper format

[17] The standard uses "in-laboratory" and "between-laboratory" to distinguish repeatability from reproducibility. See also ISO 21748: *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.*

[18] 2.83 is the Student's T value for 5 repeated measurements (4 degrees of freedom) at the two sigma level.

a. GUM Terminology

In addition to standard measurement, mathematical, and statistical terms and methods, the GUM uses terminology specifically developed for its method. The following are key concepts and terms.

Standard Uncertainty: In the GUM, all sources of measurement uncertainty are treated as if they are standard deviations of probability distributions. Standard uncertainty is the uncertainty of a measurement expressed as a standard deviation.

Input Quantities: are the quantities that determine the measured value, sometimes called the "output quantity".

Influence Quantities: are parameters that affect the input quantities and, through them, the measurement result.

> The area of a square piece of material is calculated from **two input quantities**, length and width. These quantities may be affected by **influence quantities** such as temperature and the resolution of the measuring instrument.

Type A Evaluation (of uncertainty): is an evaluation of uncertainty by the statistical analysis of a series of observations. (Type A uncertainties are not random errors.)

Type B Evaluation (of uncertainty): is an evaluation of uncertainty by means other than the statistical analysis of series of observations. (Type B uncertainties are not systematic errors.)

Combined Standard Uncertainty: is calculated by squaring all the significant Type A and Type B uncertainties, adding them together, and then taking the square root of the sum[19]. This is sometimes called the "root-sum method".

Expanded Uncertainty: is the combined standard uncertainty multiplied by a **coverage factor, k.** The expanded uncertainty defines an *interval* around the measured value. The value of the measurand is expected to be within this interval to an established confidence level, usually 95%.

Each of these terms will be illustrated and described more thoroughly later in this guide.

b. GUM Measurement Description

A measurement is considered to be a function of the all the *input quantities* that affect the measurement. Sometimes, as in the tensile strength example, the function is a known equation. For many tests, however, the "function" is not well defined. The GUM assumes the *measurement result, y,* is caused by one or more input quantities, which are designated $x_1, x_2, .., x_n$ acting through some functional relationship, f:

$$y = f(x_1, x_2, ..., x_n).$$

Returning to the tensile strength at break example, the measurement model is equation 3, which could be written to explicitly show the input quantities as

$$S = f(F,T,W) = F/TW$$

The input quantities are force (load), thickness, and width. In the sections following, we will use the GUM process to evaluate the influence quantities that affect them, the uncertainty contributions that result, and the method by which they should be combined.

---

[19] This is a simplified definition that applies in many test situations. In cases where influence quantities are correlated the calculation is more complicated and the GUM should be consulted.

The GUM assumes that the uncertainty in the measurement result can be calculated by combining the uncertainties of the input quantities. Uncertainties in input quantities may result from more than one source or influence. Among the influence quantities that may affect measurement uncertainty are the following:[20]

- Repeatability
- Resolution
- Reproducibility
- Reference Standard Uncertainty
- Reference Standard Stability
- Environmental Factors
- Measurement specific contributors
  - Alignment, scale, evaporation, mismatch, etc.
- Contributions required by method
  - ASTM, ISO/IEC, Military Procedure, etc.
- Accreditation requirements

If practical, input quantities should be varied to determine their effects on the measurement result. An uncertainty estimate should be based, as much as possible, on experimental data. If available, check standards, control charts, or other measurement assurance methods should be used to establish that a measurement system is in statistical control.

One of the characteristic features of the GUM is its designation of all uncertainty contributors as Type A or Type B. There are no other categories. Type A uncertainty estimates are derived from the statistical analysis of test data[21]. Any uncertainty contributor that is not derived from statistical analysis of test data is a Type B uncertainty contributor[22]. Type A and Type B uncertainty contributions, once determined, are both "standard uncertainties".

c. Type A Evaluation of Standard Uncertainty

Type A uncertainties are based upon repeated measurements from a controlled process and are described by the familiar Normal (or "Standard") probability distribution that yields an average and standard deviation for the set.

The formulas for average and standard deviation have already been listed (equations 1 and 2).

The GUM uses the term **standard uncertainty,** *s*, for the standard deviation of measurement results. The same statistic is also frequently called "experimental standard uncertainty".

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

Taking more samples generally improves estimates of the average and standard uncertainty. The statistic, **degrees of freedom**, is calculated from the number of measurements. For simple averages,

DOF= n-1.

The standard deviation for the average of a set of 30 values has 29 degrees of freedom[23].

---

[20] A2LA R205

[21] Sr in the tensile test example is a Type A uncertainty estimate.

[22] An uncertainty contributor based upon statistical analysis of another organization's measurement data is Type B. It is only Type A if the contributor is based on our own data.

[23] Degrees of freedom can be calculated for curve-fits and other cases beyond the scope of this paper.

It seems obvious that a standard uncertainty estimated based on 30 repeated measurements is likely to be better than one based on 5, such as in the tensile test example. However, large numbers of repeated tests may be expensive or otherwise impractical.

It is best to use a calculator or spreadsheet program for statistical calculations. The functions AVERAGE and STDEV in Excel can be used to find the average and standard deviation of test results quickly and easily. In Excel the standard deviation of the average can be calculated as STDEV/SQRT(COUNT).

d. Type B Evaluation of Standard Uncertainty

Type A uncertainty estimates apply standard statistical methods to test data based on the assumption of a normal probability distribution.

However, many factors that are not based on repeated measurements may contribute to uncertainty of measurement. Values from reference books, manufacturer's specifications, ASTM standards, experience, and many other sources of uncertainty are included in this category that the GUM calls "Type B".

The question of how to combine statistical and non-statistical uncertainty contributors had vexed the measurement community for many years. The GUM provided a practical and creative answer: treat non-statistical uncertainties as if they were statistical uncertainties with standard deviations. To do this, probability distribution functions were developed for common non-statistical uncertainty distributions, and the necessary formulas were developed for calculating standard uncertainties for each distribution[24].

Type B uncertainties are assumed to have infinite degrees of freedom[25] because they are not improved by additional repeated measurements.

The most common distributions and the formulas for their standard uncertainties follow.

e. Distribution

It might seem surprising that the normal distribution would be a Type B non-statistical uncertainty distribution. However, its use is very common and in calibration reports, national and international standards, test procedures, manufacturer's manuals, and many other technical documents.

Uncertainties based on normal distributions are Type B if they are not the result of our own measurement data. (In the example, the S_r calculated from the five tensile tests is Type A. The S_R values listed in ASTM D638 tables are Type B.)

For a standard uncertainty or standard deviation value of ±a reported at a coverage or multiplication factor of k,[26] the standard uncertainty is

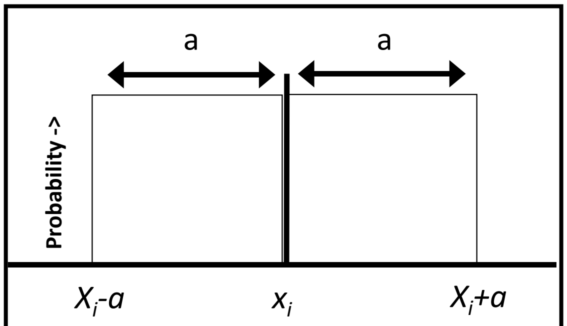
$$u_N = a / k .$$

Returning to the tensile test example, the ASTM standard requires that the load indicator have an accuracy of 1% or better. Assume we have a calibration certificate that shows an expanded uncertainty at the time of calibration

[24] The methods used to determine Type B distribution functions are contained in the GUM but are beyond the scope of this introduction.
From M3003 --
[25] For this reason, it is very important to set the interval large enough to be certain that there is a negligible probability that the uncertainty could be greater.
[26] A coverage factor of 1 corresponds to 68% containment, k=2 corresponds to 95%, and so on. These are the same statistics as for standard deviations of a normal population.

of ±1% of indication, evaluated with a coverage factor of k=2, at a 95% confidence level. For our average measured force value of 852 lb, ±1% is ±8.5 lb, which, by the equation above, gives us a standard uncertainty of

$$u_N = \frac{8.5lb}{2} = 4.25lb.$$



X_i, the expected value, lies in the middle of an interval of range 2a.

Figure 2

Rectangular Distribution

The rectangular, or "continuous", is another very common distribution for Type B uncertainties. It applies to tolerances, specifications, and reference book values, among many other parameters. It is also the default distribution to be used whenever the actual distribution is not known.

The equation for the standard uncertainty of a continuous distribution of equal values between the limits +a to –a, the standard uncertainty is

$$u_R = \frac{a}{\sqrt{3}}. \qquad 11)$$

The force indicator on our tensile test stand has a resolution of 1.0 lb. This resolution can be evaluated as a rectangular distribution with containment limits of ±0.5 lb. This gives a standard uncertainty of

$$u_{R1} = \frac{0.5lb}{\sqrt{3}} = 0.29lb. \qquad 12)$$

In order to verify that the specimen thickness and width complied with the standard requirement of ±0.02 in, we measured and recorded the dimensions of the test specimens to the nearest 0.001 in. This uncertainty is described by another rectangular distribution:

$$u_{R2} = \frac{0.0005in}{\sqrt{3}} = 0.00028in. \qquad 13)$$

This uncertainty applies to both Width and Thickness input quantities.

Triangular Distribution

It may be the case that we know that there is a tendency for the values of an uncertainty contributor to be near the center of an interval. For example, imagine two test specimens lying side by side on an aluminum plate. Assume the room temperature is controlled with a control limit of ±a. After the specimens have reached thermal equilibrium, the most likely value for the difference in temperature between them is zero.

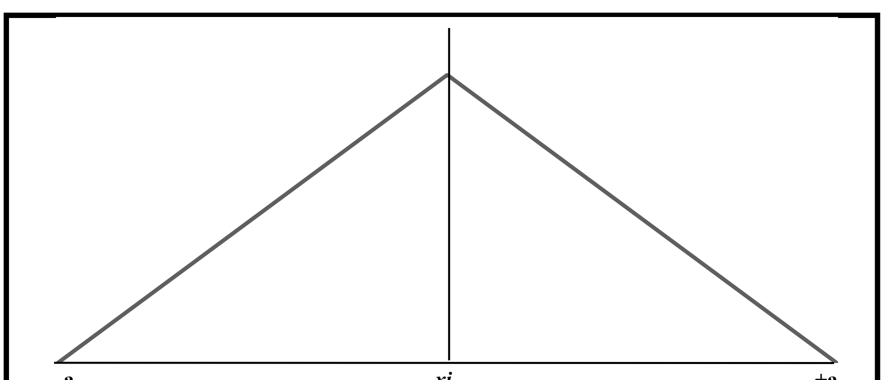


Figure 3. The triangular distribution is used to model cases where containment limits are known and values are more likely to be near the mean than at the extremes.

The triangular distribution may be used in such a case. A rectangular distribution may be used instead, but a slightly larger estimated uncertainty will result.

$$u_T = \frac{a}{\sqrt{6}}.$$

This distribution does not apply to the tensile strength example.

U Distribution

This distribution models situations where the most likely value of a measurand is at or near the containment limits. For example, because of the way thermostats work, room temperature tends to be near the maximum allowed deviation from the set point, i.e., the room temperature is most likely to be too hot or too cold relative to the set point. Applications of the U-distribution are also common in microwave and RF testing. The equation for the standard uncertainty for this distribution is

$$u_U = \frac{a}{\sqrt{2}}.$$

This distribution does not apply to the tensile strength example.

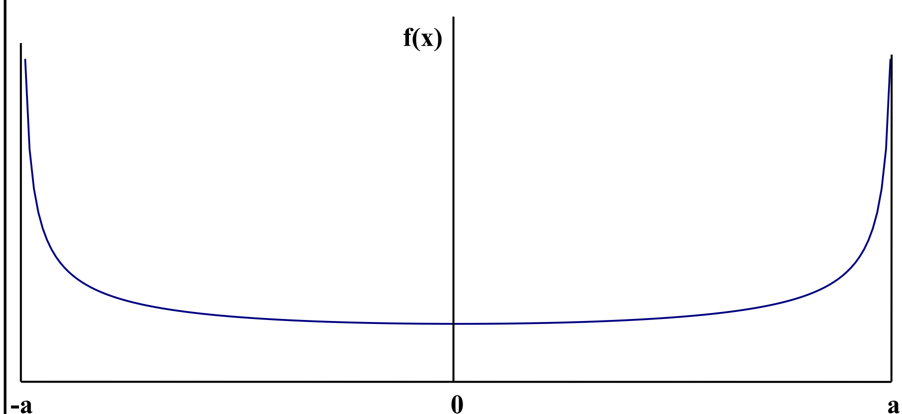


Figure 4. The U distribution models cases where the value of a measurand is likely to be near the containment limits.

f. Type B Summary

For containment limits $\pm a$, the standard uncertainty estimates associated with the various Type B probability distributions described in this guide are as follows:

Rectangular: $\frac{a}{\sqrt{3}} \cong 0.5774a$

Triangular: $\frac{a}{\sqrt{6}} \cong 0.4082a$

U-Shaped: $\frac{a}{\sqrt{2}} \cong 0.7071a$

g. Uncertainty Normal $u = a / k$, when k = coverage factor **Budget**

A table listing the sources and values of uncertainty components is an uncertainty budget. It is a useful tool but there is no mandated format and many forms are used. It is quite possible to evaluate and combine uncertainty contributions without using a budget. Along with the budget table or other listing of the constituent uncertainties, it is important that a well-documented narrative be available for every uncertainty analysis. An independent, detailed exposition is not needed every time an analysis is undertaken, however. If the conditions and assumptions used to estimate an uncertainty are the same in one case as they were in a past case, then the narrative developed for a past case is applicable to the present case and need not be duplicated. However, it may be necessary to update the values of specific uncertainty contributors if new information becomes available.

For illustration, the tensile test example uncertainties calculated thus far are listed in the table below.

Table 3. Tensile Strength Test Uncertainty Budget				
Uncertainty Source	**Standard Uncertainty**	**Type**	**Units**	**Distribution**
Repeatability (S_r)	482.8	Type A	PSI	Normal
Resolution (Thickness)	289	Type B	µin	Rectangular
Resolution (Width)	289	Type B	µin	Rectangular
Reference Standard (Force Gage) Resolution	0.29	Type B	lb	Rectangular
Ref. Standard Calibration Certificate	4.25	Type B	lb	Normal

Correlated Input Quantities

Once the uncertainty contributions from all significant influence quantities have been determined, the next step in the GUM 8-step process is to identify covariances and correlations, if any, in the input quantities.

Correlation occurs when the values of input quantities are not independent. For example, in the tensile strength measurement we've been examining, the measurements of the thickness and width of the test specimen would be correlated if both quantities were measured with the same device. Correlated quantities do not combine through least square summation.

> **Example:** Four 100 pound weights are used together to calibrate a load cell at 400 pounds. These weights were all calibrated at the same laboratory on the same scale. The uncertainty of each weight is said to be 100 mg. At least some of this uncertainty will be a bias. For the sake of example, assume the entire 100 mg is a positive bias from the expected value. The resultant bias caused by using the four weights together is 400 mg. However, if the method of least square summation were used the result would only 200 mg[27], which is only half the true uncertainty.

Correlated input quantities are common in testing and a simple method for addressing them is to add the correlated uncertainties together and use that sum in the combined uncertainty calculation. This method is conservative but may result in a much larger total uncertainty estimate than would be obtained by a more rigorous approach. In the table below, resolution uncertainties for thickness and width are added together because they are correlated.

For more information, consult the GUM on the topic.

Table 4.				
Uncertainty Source	**Standard Uncertainty**	**Type**	**Units**	**Distribution**
Repeatability (S_r)	482.8	Type A	lb	Normal
Thickness & Width Correlated (289+289)	578	Type B	µin	Rectangular
Reference Standard (Force Gage) Resolution	0.29	Type B	lb	Rectangular
Ref. Standard Calibration Certificate	4.25	Type B	lb	Normal

[27] $\sqrt{100^2 + 100^2 + 100^{2} + 100^2} = 200$

Calculating the Measured Value to Report

The next step of the GUM 8-step process is to calculate the quantity value to be reported for the measurement, usually an average measured or calculated value. The ASTM D638 method requires a calculation of maximum tensile strength which we have already done. Here we will also calculate the average tensile strength value.

> Using the tensile strength formula with average values for force, thickness, and width, we calculate the average value:
>
> $$S = \frac{F}{TW} = \frac{852lb}{(0.125in \times 0.500in)} = 13632\ lb/in^2$$

Sensitivity Coefficients

Before the uncertainty contributions from the input quantities can be combined, they must all be in the same units. (You cannot add apples and oranges, or inches and millimeters.) Returning to our tensile strength example, notice that the uncertainty for Thickness and Width is in units of "microinches" and the uncertainty for Force is in "pounds", but Tensile Strength is given in PSI[28]. Before the uncertainties in all the input quantities can be combined, they must be converted into the units associated with the measured value for the test, tensile strength at break. The applicable units are PSI.

Sensitivity Coefficients is the GUM term for conversion factors that convert from input quantity units into units of the measurand. These conversions may be made at any time in the uncertainty estimation process, but they must be performed before the uncertainties are combined. The conversion is often quite simple. For example, multiplying micro-inch values by the sensitivity coefficient 1,000,000 converts them into inches. Unfortunately, determining sensitivity coefficient values is sometimes a difficult process.

When the model function is known, sensitivity coefficients can be readily determined using calculus. (An approximation method that does not require calculus is given in Appendix C.)

Mathematically, sensitivity coefficients are partial derivatives of the model function *f* with respect to the input quantities. In particular, the sensitivity coefficient c_i of the input quantity x_i is given by

$$c_i = \frac{\partial f}{\partial x_i} \qquad 20)$$

Example: Sensitivity Coefficients Using Calculus

The model function we're using for the tensile strength determination is

$$S = \frac{F}{TW}$$

where *S* is the tensile strength, *F* is the load needed to break a test specimen, and *T* and *W* are the thickness and width respectively of the test specimen. We obtain the sensitivity coefficients as follows:

$$c_F = \frac{\partial S}{\partial F} = \frac{1}{TW} = \frac{S}{F};\ c_T = \frac{\partial S}{\partial T} = \frac{-F}{T^2W} = \frac{-S}{T};\ c_W = \frac{\partial S}{\partial W} = \frac{-F}{TW^2} = \frac{-S}{W}\ .$$

The average thickness of the test specimen is 0.125 *in*, the average width is 0.500 *in*, the average force needed to break the specimens is 852.0 *lb* and the average tensile strength is 13632 lb/in^2. With these values we can determine the values of each of the sensitivity coefficients:

---

[28] lb/in^2

$$c_F = \frac{S}{F} = \frac{13632 \, {}^{lb}\!/_{in^2}}{852 lb} = +16.0 in^{-2};$$

$$c_T = \frac{-S}{T} = \frac{-13632 \, {}^{lb}\!/_{in^2}}{0.125 in} = -109056 \frac{lb}{in^3};$$

$$c_W = \frac{-S}{W} = \frac{-13632 \, {}^{lb}\!/_{in^2}}{0.500 in} = -27264 \frac{lb}{in^3}.$$

Determinations of sensitivity coefficients must take place at or very close to the input quantity values actually measured in a test.

In the table below, the uncertainty contributors are listed in base units, Type A or Type B is indicated, along with the distribution assumed (N for normal, R for rectangular). Multiplication by the Sensitivity Coefficient gives the uncertainty in measurement units of PSI. The table also lists degrees of freedom for each standard uncertainty.

Table 5.						
Uncertainty Source	**Standard Uncertainty**	**Units**	**Type, Dist**	**Sensitivity Coefficient**	**Standard Uncertainty (PSI)**	**DOF**
Repeatability (S_r)	482.8	PSI	A, N	1	483	4
Micrometer Resolution (Thickness)	0.0000289	Inch	B, R	109056	31.5	∞
Micrometer Resolution (Width)	0.0000289	Inch	B, R	27264	7.88	∞
Reference Standard (Force Gage) Resolution	0.289	Lb force	B, R	16	4.62	∞
Ref. Standard Calibration Certificate	4.25	Lb force	B, N	16	68.0	∞

Combining the Contributors

Once all of the values of the standard uncertainty contributors u_i have been estimated and the sensitivity coefficients c_i have been determined and applied, they are combined by "root-sum-square", i.e., taking the square root of the sum of the squares of the uncertainty estimates in order to determine the combined standard uncertainty.

$$u_c = \sqrt{\sum_i c_i^2 u_i^2} \, .$$

Because the uncertainties are combined by root sum square, it is common practice to list them in absolute value rather than show negative signs in the budget table.

Table 6.

Index	Uncertainty Source	Standard Uncertainty	Units	Type, Dist	Sensitivity Coeff	Standard Uncertainty (PSI)	DOF
1	Repeatability (S_r)	482.8	PSI	A, N	1	483	4
2	Resolution (Thickness)	0.0000289	Inch	B, R	109056	31.5+7.88= 39.4	∞
	Resolution (Width)	0.0000289	Inch	B, R	27264		∞
3	Reference Standard (Force Gage) Resolution	0.289	Lb force	B, R	16	4.62	∞
4	Ref. Standard Calibration Certificate	4.25	Lb force	B, N	16	68.0	∞
	u_c, Combined Uncertainty (RSS):					489	PSI

It is not necessary to use an uncertainty budget for these calculations and there is no mandatory format for an uncertainty budget, if one is used. Most people find the uncertainty budget an efficient and practical way to manage the calculations.

Notice the term near the middle where 31.5 and 7.88 PSI terms are combined prior to squaring. This is a conservative method to account for correlation. Correlation effects can be quite large in some situations. In this example, the effect is really negligible and only included for educational purposes. The reader is referred to the GUM where this matter is addressed in depth.

Calculating the Expanded Uncertainty

The expanded uncertainty *U* is obtained by multiplying the combined standard uncertainty by coverage factor *k*:

$$U = ku_c.$$

The procedure for determining the coverage factor is presented below. The reader is urged to consult the GUM for more information and the rationale behind the procedure.

Estimating the Coverage Factor

Multiples of the standard deviation of a population characterized by a normal probability distribution provide the probabilities that a value lies within a specified range. In the same way, the coverage factor provides the multiplier to be applied to the combined standard uncertainty to ensure that the measured value lies within the uncertainty range to some specified confidence level. K=1 provides 68% confidence, K=2 provides 95% confidence, K=3 provides 99% confidence, and so on. The coverage factor is a function of the effective degrees of freedom for the combined uncertainty.

K=2 is commonly used for calibration and test reports. This value is appropriate when Type B uncertainty components dominate the uncertainty budget or when Type A components have been established with 30 repeated measurements or more.

For other situations, such as the tensile strength test example, where the Type A uncertainties dominate and where there are fewer than 30 degrees of freedom, a larger value for k is required[29]. The appropriate k value is based on the Student's T distribution (Appendix 2) and the effective degrees of freedom for the measurement. The reader is encouraged to refer to the GUM sections dealing with coverage factor and degrees of freedom.

Degrees of freedom can be conservatively estimated by assuming infinite degrees of freedom whenever Type B components comprise more than half the combined uncertainty and using the actual degrees of freedom from the Type A portion of the budget otherwise. In the case of tensile testing, degrees of freedom=4.

A rigorous estimate can be made by following the procedure below.

1) Obtain the estimate of the measurand y and the estimate of the combined standard uncertainty $u_c(y)$.

2) Estimate the effective degrees of freedom ν_{eff} using the Welch-Satterthwaite formula:

$$\nu_{eff} = \frac{u_c^4(y)}{\sum_{i=1}^{n} \frac{c_i^4 u^4(x_i)}{\nu_i}}$$

where ν_i is the degrees of freedom of the estimate[30] of the magnitude of the uncertainty contributor x_i.

This equation is complicated but can be readily performed with a spreadsheet.

We can calculate the effective degrees of freedom ν_{eff} for the tensile test example as follows:

Because we have already applied the sensitivity coefficients, the terms $c_i u(x_i)$ have already been calculated in PSI and are u_1=483 with ν_1=4, u_2=39.3 with $\nu_2=\infty$, u_3=4.62 with $\nu_3=\infty$, and u_4=68.0 with $\nu_4=\infty$. u_c= 489 PSI.

$$\nu_{eff} = \frac{(489)^4}{\frac{(483)^4}{4} + \frac{(39.3)^4}{\infty} + \frac{(4.62)^4}{\infty} + \frac{(68.0)^4}{\infty}} = 4 .$$

Consulting the Student's *t*-table, we find that the value of *t* corresponding to 4 degrees of freedom at the 95% level of confidence is t = 2.8. So the expanded uncertainty of the tensile strength result is

$$U = ku_c = (2.8)(489\frac{lb}{in^2}) = 1{,}369 PSI .$$

This is about 10% of our test result of 13600 psi. We round the final uncertainty estimate to no more than two significant figures (U=1,400 PSI) so as not to convey the impression of greater accuracy than is warranted. Standard rounding practice such as the one found in section 6.4 of ASTM E29 should be followed, although it is common practice always to round uncertainty estimates to one or two significant figures. It is best to do these

---

[29] It should be noted that the ASTM standard provided the appropriate K factor for the uncertainty due to repeatability: 2.83.
[30] The degrees of freedom of a Type A evaluation based on n repeated measurements is simply $\nu = n - 1$. If n independent observations are used to determine both the slope and intercept of a straight line by the least squares method, then the degrees of freedom of their respective standard uncertainties is $\nu = n - 2$. In general, for a least-squares fit of m parameters to n data points the degrees of freedom of the standard uncertainty of each parameter is $\nu = n - m$.

uncertainty calculations with a spreadsheet so that intermediate-rounding errors can be avoided. In this case, rounding errors were negligible.

If we had decided, because Type A dominates the uncertainty budget, to just use the k value associated with the Type A degrees of freedom for the expanded uncertainty, we would have gotten the same result without making complicated calculations.

In some tests there may be multiple, significant Type A uncertainties with differing degrees of freedom. For such a case, the Welch-Satterthwaite is useful.

Reasonability

In the end, every uncertainty estimate should be subjected to a reasonability check. The analyst should ask questions such as “Is this estimate in line with what I know about the nature of the measurement and of the material?” “Can this estimate be supported with proficiency testing data, or data accumulated as part of a measurement assurance program?” Uncertainty estimates that look strange -- either too big or too small -- should be re-evaluated, looking first for mathematical blunders, second for uncertainty contributors whose magnitudes may have been poorly estimated or neglected. Finally, it may be necessary to revise the mathematical model.

Human judgment based on sound technical experience and professional integrity is of paramount importance in evaluating uncertainty.

Reporting Uncertainty

When reporting the result of a measurement, at a minimum one should provide the following:

1) A full description of how the measurand Y is defined;

2) The result of the measurement as $Y = y$ ⊚ U and give the units of y and U;

3) The value of the coverage factor k used to obtain U;

4) The approximate level of confidence associated with the interval y ⊚ U and state how it was determined.

The numerical values of the estimates of the measurand and expanded uncertainty should not be given with an excessive number of significant digits. Uncertainty estimates should be quoted to no more than two significant figures.

When stating measurement results and uncertainty estimates it is always advisable to err on the side of providing too much information rather than too little and this information must be stated as clearly as possible.

The statement of our tensile strength result might take the following form:

“The maximum tensile strength at break is determined to be 14,400 PSI based on tests of five specimens conducted in accordance with ASTM D638.

“For these test specimens, the average tensile strength at break was also determined. Average tensile strength at break = 13,636 PSI ⊚ 1,400 PSI. The uncertainty listed is the expanded uncertainty based on a coverage factor of 2.8 (95% confidence) calculated in accordance with the GUM.”

It should be noted that we did not improve the ASTM D638 test by performing uncertainty calculations. We also did not eliminate the unknown bias mentioned earlier that is associated with this measurement. We did provide additional information that might be of value and, in the process, demonstrated the GUM method with an actual test example.

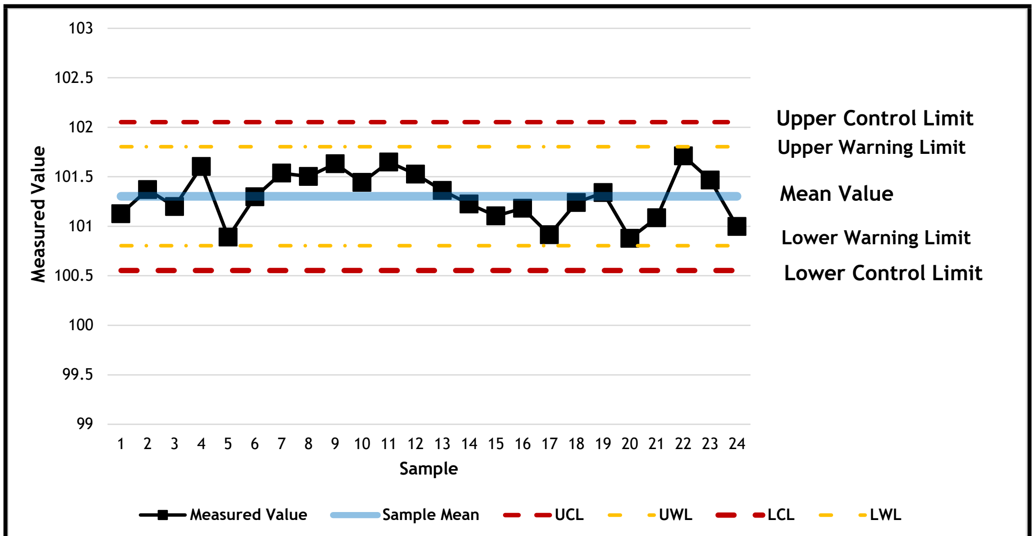
END

APPENDIX A

Control Charts

A control chart[31] (also called "Shewhart Chart") is a plot of some characteristic statistic for test data with a center line that is the mean value, and upper and lower lines at established control limits. There may also be upper and lower warning limits shown on the chart.

In the figure below, the characteristic statistic is a measured value.



A Control Chart

It is common practice to set upper and lower control limits on a control chart at three times the standard deviation ("3 ◎") to ensure that approximately 99.7% of measurements are within the limits for a process in statistical control.

It is sometimes useful to plot 2 ◎ upper and lower warning limits which contain 95.0% of the measurements for a measurement process in statistical control.

Control charts can be used to estimate measurement uncertainty if the following conditions are met:

1) The control test sample has a certified or otherwise known or accepted value. Then, bias in the measurement process may be identified and corrected in the calculation of measurement results. There will be some uncertainty associated with bias corrections, so it may be necessary to identify and quantify this uncertainty and root-sum-square it with the standard deviation associated with the control limits.

2) The value of the measurand represented by the control sample should be close to the value of the measurand actually obtained during routine testing since, in general, the uncertainty of a measurement will be some function of the "level of the test", or value of the measurand. Consequently, it may be necessary to track several control samples at different measurement levels to properly assess the measurement uncertainty across the range of the measurand encountered in the testing laboratory.

---

[31] See also ASTM E2554: *Standard Practice for Estimating and Monitoring the Uncertainty of Test Results of a Test Method in a Single Laboratory Using a Control Sample Program* and ISO 21748: *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertain estimation.*

3) The measurement process for control samples should be the same as for routine samples, including subsampling and sample preparation. If it is not, then additional uncertainty components may have to be considered.

4) The measurement process must be in statistical control as demonstrated by the control chart. This means that a sufficient number of data points must be collected to establish that the process in in control and to ensure that the estimate of the population standard deviation is reasonably accurate. There are no universally applicable rules, but 20 – 25 subgroups of 4 or 5 are generally considered adequate for providing preliminary estimates. Measurement processes that are not in statistical control must be brought into control before the control chart can be properly constructed.

Control charting is often a reliable, simple tool for estimating measurement uncertainty for testing.

However, control charts are not practical in all testing situations. Tests that are conducted infrequently, for which multiple repeated measurements are not practical, or for which a reference material is not available, do not lend themselves to control charting.

APPENDIX B

Student's t-Table

Degrees of Freedom	One Sigma, 68%	Two Sigma, 95%	Three Sigma, 99.7%
1	1.818993	12.7062	212.205
2	1.311578	4.302653	18.21631
3	1.188929	3.182446	8.891456
4	1.134397	2.776445	6.434848
5	1.103668	2.570582	5.376025
6	1.083976	2.446912	4.800243
7	1.070287	2.364624	4.442125
8	1.060224	2.306004	4.199149
9	1.052515	2.262157	4.023987
10	1.046423	2.228139	3.891955
11	1.041486	2.200985	3.788982
12	1.037405	2.178813	3.706487
13	1.033976	2.160369	3.638947
14	1.031053	2.144787	3.582653
15	1.028533	2.13145	3.535025
16	1.026337	2.119905	3.494212
17	1.024407	2.109816	3.458854
18	1.022698	2.100922	3.42793
19	1.021172	2.093024	3.400658
20	1.019804	2.085963	3.376428
21	1.018568	2.079614	3.354759
22	1.017447	2.073873	3.335267
23	1.016426	2.068658	3.317639
24	1.015492	2.063899	3.301622
25	1.014634	2.059539	3.287005

Degrees of Freedom	One Sigma, 68%	Two Sigma, 95%	Three Sigma, 99.7%
26	1.013843	2.055529	3.273611
27	1.013112	2.051831	3.261294
28	1.012434	2.048407	3.249929
29	1.011804	2.04523	3.23941
30	1.011216	2.042272	3.229646
31	1.010667	2.039513	3.220559
32	1.010152	2.036933	3.21208
33	1.009669	2.034515	3.204151
34	1.009215	2.032245	3.19672
35	1.008788	2.030108	3.189741
36	1.008384	2.028094	3.183175
37	1.008003	2.026192	3.176986
38	1.007642	2.024394	3.171142
39	1.007299	2.022691	3.165615
40	1.006974	2.021075	3.160381
50	1.004446	2.008559	3.120076
60	1.002768	2.000298	3.093713
70	1.001572	1.994437	3.075127
80	1.000677	1.990063	3.061319
90	0.999983	1.986675	3.050657
100	0.999427	1.983972	3.042175
200	0.996936	1.971896	3.004535
300	0.996109	1.967903	2.992177
400	0.995696	1.965912	2.986032
500	0.995448	1.96472	2.982357
∞	0.994458	1.959964	2.967739

APPENDIX C

Case 1: Estimating Sensitivity Coefficients Mathematically

The model function we're using for the tensile strength determination is

$$S = \frac{F}{TW}$$

The average thickness of the test specimen is 0.125 *in*, the average width is 0.500 *in*, the average force needed to break the specimens is 852.0 *lb* and the average tensile strength is 13632 *lb/in*^{2}. With these values we can estimate the values of each of the sensitivity coefficients.

The method is a mathematical approximation based on the model function. It consists of making a calculation using actual data, changing one of the quantities of interest, Force, for example, by a small and arbitrary amount and then recalculating the output quantity. From the change in output quantity value caused by a small change in one input quantity, the sensitivity coefficient can be directly calculated.

Below, this method will be used to estimate each sensitivity coefficient for our tensile strength example. One parameter will be changed for each calculation while the others will be held constant.

Using the measured load of 852 lbs, thickness of 0.125 in, and width of 0.500 in in the formula above gives a value of S = 13632 PSI.

If we hold everything else constant and increase F by about 1% to 860 lbs, which we call F', we can calculate what would be the changed tensile strength at break, S'=860/(0.125×0.500) = 13,760 PSI.

The difference in tensile strength is S'-S= ◎S = 13760-13632 = 128 PSI

The difference in force is F'-F= ◎F= 8 lbs

With this information, we calculate the sensitivity coefficient for force, $c_F \approx$ ◎S/◎F = 128/8 = 16 PSI/lb

The amount of the assumed change is not important, but it must be small.

Similarly, if we make an imaginary increase in T of about 1% to 0.130 in., the tensile strength at break would be S'=852/(0.130×0.500) = 13,108 PSI

The difference ◎S = 13108-13632 = 524 PSI

The difference ◎T= 0.005 in. c_T c
$c_T \approx$ ◎S/◎T = 524/0.005 = 104,800 PSI/in

If we make an imaginary increase in W of about 1% to 0.505 in., the tensile strength at break would be S'=852/(0.125×0.505) = 13,497 PSI

The difference ◎S = 13497-13632 = 135 PSI
The difference ◎W= 0.005 in.
$C_W \approx$ ◎S/◎W = 524/0.005 = 27,000 PSI/in

Comparison to the sensitivity coefficients calculated using calculus will show close but not perfect agreement. These estimates are close enough, however, to be used.

Case 2: Estimating Sensitivity Coefficients Experimentally

Experimental determination of sensitivity coefficients is preferable to approximation, if practical[32].

Sensitivity coefficients may be determined experimentally through a process very similar to that described above for mathematical estimation. The principal difference is that, in this case, the small changes in input quantities used to calculate sensitivity are real, based on actual data from tests.

For this case, we will use the same tensile strength example and model function:

$$S = \frac{F}{TW}$$

Let us identify the set of test specimens used in our examples up to now as Set 1. We select another set of 5 test specimens, Set 2, that are slightly thicker[33] and repeat the test. The results for the Set 1 and Set 2 tests are shown in the table below.

Set	Average Measured Thickness, T	Average Measured Width, W	Average Area, TW	Average Measured Load, F	Average Calculated S
1	0.125 in	0.500 in	0.0625	852.0 lb	13632 PSI
2	0.128 in	0.500 in	0.0640	890.6 lb	13916 PSI

The difference in S between the two sets is ΔS = 13916-13632 = 284 PSI.
The difference in thickness between the two sets is ΔT= 0.128 – 0.125 = 0.003 in.
The sensitivity coefficient for thickness, c_T = ΔS/ΔT = 284/0.003 = 94,667 PSI/in.

The same process could be used again to estimate the coefficient for width, if an appropriate set of test specimens could be located. This set of specimens would have the same average thickness but different average width than the specimens in the previous set.

Unfortunately, experimental determination of sensitivity coefficients is not always practical and sometimes not possible. Multiple sets of test specimens with the necessary dimensions might not be available. The time and expense to perform multiple tests for the purpose of estimating sensitivity coefficients might be unacceptable expenditures. The precision of the indicator might not be adequate to reveal the effect of small changes in input quantities. There might not be enough tolerance allowable in the input quantity to perform experimental determinations for sensitivity.

Sometimes, a sensitivity coefficient can be determined by an experiment as simple as changing room temperature by one degree and observing the corresponding change in some quantity.

[32] The model function may not include all influence quantities; it may be an approximation of some more complex function, it may suffer from other limitations. The response of the model function is necessarily hypothetical. Experimental data, on the contrary, reveals the actual performance of the test system when parameters are varied.
[33] The new specimens must remain within the ASTM D638 tolerances.

References

"Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025", ILAC-G17

"Standard Test Method for Tensile Properties of Plastics", ASTM 638

BIPM JCGM 100:2008, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM 1995 with minor corrections)*.

BIPM JCGM 200:2012, *International vocabulary of metrology - Basic and general concepts and associated (VIM) 3rd edition (2008 version with minor corrections).*

"P103A2LA Policy on Measurement Uncertainty for Testing Laboratories", American Association for Laboratory Accreditation, 2013

DOCUMENT REVISION HISTORY

Date	Description
01/05/19	➢ Integrated into Qualtrax
09/22/19	➢ Updated Header/Footer to current version ➢ Updated format and font for consistency