

# Applied Statistics in Chemistry

The latest version of this document is available from [www.consol.ca](http://www.consol.ca) (Teaching link).

The fundamental hypothesis in statistics is the *Null Hypothesis*. The null hypothesis states that random error is sufficient to explain differences between two values. Statistical tests are designed to test the null hypothesis. Passing a statistical test means that the null hypothesis is retained: there is insufficient evidence to show that there is a difference between the samples.

*It is impossible to show that two values are the same; it is only possible to show they are different.*

## Significant Figures

Some values are known or defined to be exact. For example:

- the  $\frac{1}{2}$  and 2 in  $E_K = \frac{1}{2} m v^2$
- the stoichiometric coefficients and molecular formulae in chemical reactions such as
$$\text{C}_3\text{H}_8 + 5\text{O}_2 \longrightarrow 3\text{CO}_2 + 4\text{H}_2\text{O}$$
- the speed of light in a vacuum,  $c$ , is defined as  $2.99792458 \cdot 10^8$  m/s

*There is error in every observation.* Error arises due to limitations in the measuring device (ruler, pH meter, balance, etc.) and problems with equipment or methodology. The former are ‘indeterminate’ or ‘random’ errors and cannot be eliminated. Random errors limit the precision with which the final value can be reported. The latter are ‘determinant’ or ‘systematic’ errors and affect the accuracy of the final value. Analytical chemists continuously monitor for systematic errors in procedures.

### Significant figures

‘Sig-figs’ are a simple, easy to apply, quick-and-dirty method of getting approximately the correct number of decimal places in a value. The correct, but more difficult, method is to *statistically* determine the uncertainty and thus the reportable number of decimal places. This approach considers the uncertainty associated with every observation and its importance in the overall uncertainty. It is possible to gain or lose decimal places compared with the sig-figs method.

Instructors may use the term ‘sig-figs’ when they mean ‘statistically calculated number of significant digits’. This often confuses the students and instructor. Interestingly, some instructors demand the uncertainty have one sig-fig; others accept up to two; still others use a ‘3-30’ rule.<sup>1</sup> Any of these methods is acceptable as long as it is consistently applied.

To report the statistical uncertainty in the final value, the text could take the form, “Sample 123A has a lead content of  $(9.53 \pm 0.22)$  ppm at the 95 % confidence level.” The final value has the same number of decimal places as the uncertainty. (Remember the leading zero for all numbers between -1 and 1!)

### Units in calculations

Inclusion of units in calculations ensures that the final answer is not in error by a simple units conversion: joules  $\leftrightarrow$  kilojoules, grams  $\leftrightarrow$  milligrams  $\leftrightarrow$  micrograms,  $R = 8.314$  J/(mol K) = 0.08206 L atm/(mol K), etc.

Critically evaluate every answer. If you react 5 g of A with 7 g of B, is it reasonable to expect the theoretical yield be 39 g? or 240  $\mu\text{g}$ ? If you repeat a titration three times, each with 5.00 mL of the unknown, is it reasonable that the required volumes of titrant are 14.27 mL, 9.54 mL, and 9.61 mL?

## Rounding

Several rules for rounding are taught; you have probably met more than one in your courses. Everyone is adamant *their* rules are correct. The National Institute of Science and Technology (NIST) policy on rounding numbers is presented here.<sup>2</sup> (It *is* correct. ☺)

First, keep all the digits from intermediate calculations. Round the final value as follows:

If the digits to be discarded are	Round the last digit to be kept	Example
less than 5	down	3.7249999 rounded to two decimal places is 3.72.
greater than 5	up	3.7250001 rounded to two decimal places is 3.73.
exactly 5 (followed only by zeros)	even	3.72500... rounded to two decimal places is 3.72.

When manipulating data, keep all digits through intermediate calculation. Round the final value to the appropriate number of significant digits. *Don't round until the end.*

## Accuracy, Precision, and Tolerance

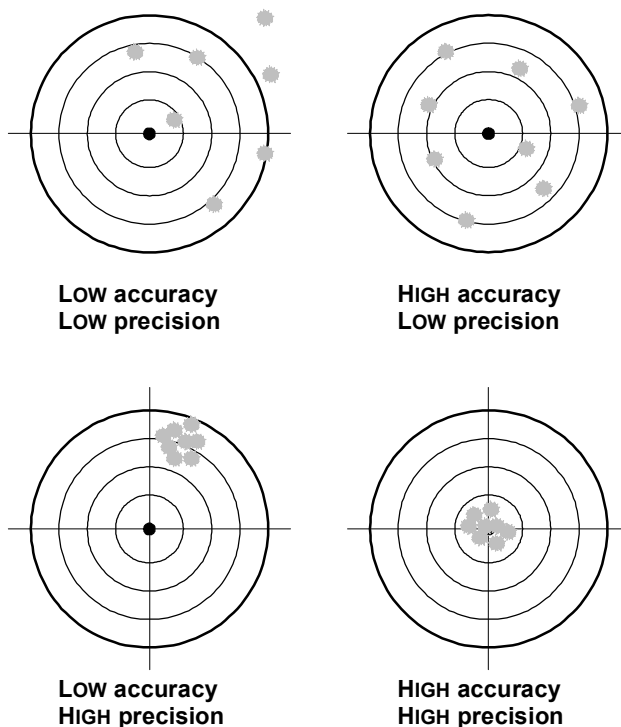
*There is no relationship between accuracy, precision, and tolerance.*

### Accuracy

Accuracy is a measure of the difference between an experimental value and the true value. Any difference is due to systematic error(s). For example, a systematic error exists if a volumetric pipet is blown out or if the edge of a ruler is used instead of the zero graduation.

Accuracy can only be determined where the 'true' value of a sample is known, i.e., a reference. Certified reference materials (CRMs) are substances that contain one or more analytes in a given matrix. They have been exhaustively characterized by several laboratories using a number of analytical techniques to provide bias-free results. CRMs are expensive! Would you pay 241\$ US for 55 g of soil containing  $432 \text{ ppm} \pm 17 \text{ ppm}$  lead at the 95 % confidence level? How about 6088\$ US for a single platinum thermocouple capable of measuring absolute temperatures to within 0.2 mK? It comes in a nice wooden box...<sup>3</sup> If no suitable CRM is available, or is too expensive, and that level of precision is not needed, an alternative is to prepare an in-house reference.

The CRM or in-house reference is used to make quality control (QC) samples. The QCs are run at the same time as the unknowns. Since their concentration is known, systematic errors can be detected by comparing the experimental value with the true value.



*Chemists who master both accuracy and precision are deadly!*

## Precision

*Every experimentally measured value has an associated uncertainty.*

Precision is characterized as the distribution of random fluctuations about the ‘true’ value. Statistics assumes that the distribution is gaussian (a.k.a. ‘normal’).<sup>4</sup> A gaussian distribution’s width is defined by a single parameter, the standard deviation,  $\sigma$ . Figure 1 illustrates the dependence on  $\sigma$ : 68.3 % of the gaussian’s area is contained between  $-\sigma$  and  $\sigma$ , 95.4 % between  $-2\sigma$  and  $2\sigma$ , and 99.7 % between  $-3\sigma$  and  $3\sigma$ . We will see that the standard deviation of a series of observations is used to determine the certainty with which we can report a value. It is *impossible* to reduce the standard deviation to zero, even with an infinite number of observations. To encompass the true value with a desired confidence, the standard deviation is multiplied by a factor,  $t$ , dependent on the number of observations and required confidence level (see *Encompassing the true value*, below).

A multitude of factors affect precision:

- instrument noise (detector sensitivity, noise, etc.)
- experimental technique (pipetting, weighing, filling, etc.)
- sample inhomogeneity

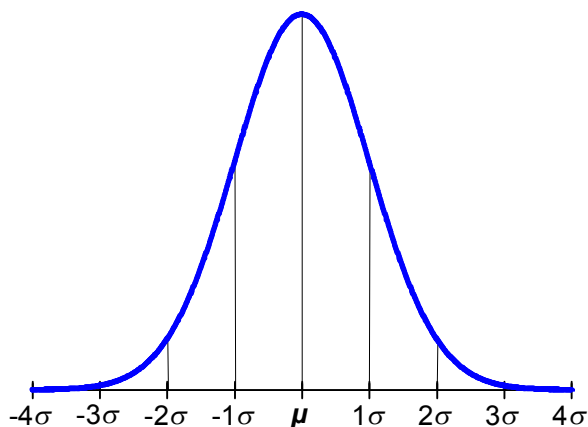


Figure 1. Gaussian distribution showing the true value,  $\mu$ , and standard deviations,  $\sigma$ .

## Tolerance

*Tolerance is not a statistical parameter*; it is the range of variation from the expected standard. For example, the tolerance of a 10.00 mL class A volumetric pipet is  $\pm 0.02$  mL. This means that the pipet is guaranteed to deliver between 9.98 mL and 10.02 mL. It *does not* mean that the pipet will deliver an average of 10.00 mL. A given pipet might routinely deliver 9.997 mL or 10.015 mL or 9.981 mL. Unlike precision, tolerance does not have a gaussian distribution. Practicing analytical chemists calibrate their pipets. Analytical chemists can repeatably deliver within  $\pm 0.002$  mL with a 10.00 mL pipet. They gain an extra decimal place and reduce the associated uncertainty by a factor of 10!

It is a systematic error if you report the volume delivered by a 10 mL pipet as  $(10.00 \pm 0.02)$  mL, which is the tolerance, when the pipet actually delivers  $(10.011 \pm 0.004)$  mL. The uncertainty in the final value will also be proportionately larger.

## Formulae and Examples

### Rejecting data (Q-test)

It is good practice to check outliers in a data set to see if they can statistically be rejected. This is done using the  $Q$ -test.

$$Q_{\text{calc}} = \frac{\text{gap}}{\text{range}} = \frac{|\text{suspect} - \text{nearest}|}{\text{largest} - \text{smallest}}$$

$Q_{\text{tab}}$  is looked up in a table and compared with  $Q_{\text{calc}}$ . If  $Q_{\text{calc}} > Q_{\text{tab}}$ , the outlier data point can be rejected at the specified confidence level. (Note: this table uses  $n$ , the number of observations; all other statistical tables use degrees of freedom.)

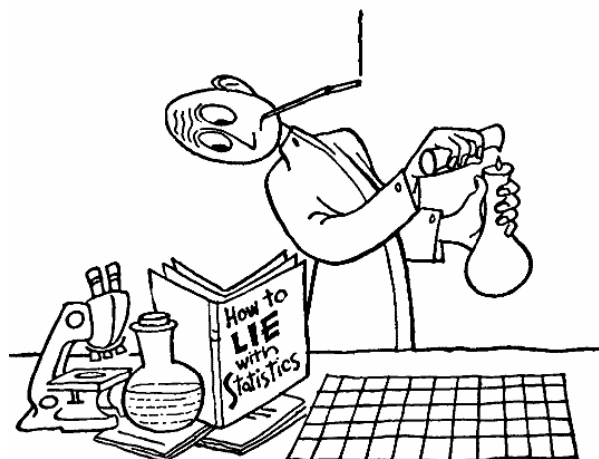
## Average

The average,  $\bar{x}$ , can be calculated as the mean, median, and mode for  $n$  observations of a sample.

The *mean* is calculated from the formula:  $\bar{x} = \frac{1}{n} \sum_i x_i$

The *median* is the middle data point after the data are sorted in ascending or descending order. If there are an even number of data points, the median is the mean of the center two data points. The *mode* is the most frequently observed value. It can only be used with large data sets — not common in analytical labs!

If the number of observations is very large (i.e., the entire population) and if no systematic errors exist, the average value becomes the true value,  $\mu$ .



(From the book of the same title.)

Often, raw data is mathematically transformed to obtain information. It is important to convert each observation to the final value before averaging. Why? Because non-linear mathematical transformations (square root, power, logarithm, etc.) skew the distribution of observations. *There is a difference* if each observation is transformed to the final value and then averaged *or* averaged and then transformed to the final value. For example, consider reading two values versus their average from a non-linear calibration curve.

## Standard deviation

The sample standard deviation,  $s$ , is a measure of the precision of a *single observation* in a series of observations. If the number of observations is very large (i.e., the entire population), the sample standard deviation becomes the population standard deviation,  $\sigma$ . Note the difference in formulae.

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} \quad \sigma = \sqrt{\frac{\sum_i (x_i - \mu)^2}{n}}$$

The standard deviation can also be viewed as the range in which we expect the next observation to be found with a certain confidence. We are often interested in the standard deviation of a value obtained from the original data, such as the *standard deviation of the average* ( $s_{\bar{x}}$ ), *slope* ( $s_m$ ), *intercept* ( $s_b$ ), etc. These are calculable from the sample standard deviation.

$$s_i = \frac{s}{\sqrt{n}}$$

The relative uncertainty (*uncertainty/average*) can be used to evaluate the precision at various points in a process or to evaluate the precision between different methods. One common calculation is percent relative standard deviation (%RSD). However, similar calculations are valid for any confidence level.

$$\%RSD = \frac{\text{standard deviation}}{\text{average}} 100\% = \frac{s}{\bar{x}} 100\%$$

### ***Encompassing the true value: confidence intervals***

The average and standard deviation can be calculated when more than one observation of a sample is made. It is not possible to determine the true value by replicate observations, but the probability of the true value being within a calculable range can be determined. Multiplying the standard deviation by a factor  $t$  (often called Student's  $t$ 's) determines the *confidence interval* (a.k.a. uncertainty and  $\Delta x$ ) of the observation at the stated confidence level.  $t$  values for different confidence levels and degrees of freedom are tabulated. Unless there is a reason to believe otherwise, the two-tailed  $t$ -value is used, which indicates that the true value could be either above or below the calculated average.

$$\mu = \bar{x} \pm \Delta x \qquad \Delta x = t s_i = \frac{t s}{\sqrt{n}}$$

**Statistics in the real world:** polls and surveys often contain a statement, "The poll/survey is accurate to within (for example) three percentage points 19 times out of 20." Statistically, this statement states that the uncertainty is  $\pm 3\%$  at the 95% confidence level. ( $19/20 \cdot 100\% = 95\%$ )

**Example:** A common analytical experiment is the gravimetric analysis of copper in brass. Five samples were analyzed and the percentage of copper in each determined to be 93.42%, 93.86%, 92.78%, 93.14%, and 93.60% by mass. The mean is 93.36% and the standard deviation is 0.417%. For five samples (four degrees of freedom) at the 95% confidence level,  $t_{(95\%, 4)} = 2.776$ . The uncertainty,  $t s / \sqrt{n}$ , is 0.52%.

The report would contain the statement, "The concentration of copper in the brass was determined to be  $(93.4 \pm 0.5)\%$  by mass at the 95% confidence level."

If a QC sample with known value ( $\mu^*$ ) is also analyzed, a  $t$ -test can be used to determine if there is a statistical difference between the experimental value and the known value. Failure of this test indicates that systematic errors may exist in the experimental method.

$$t_{\text{calc}} = \left| \frac{\sqrt{n} (\mu^* - \bar{x})}{s} \right|$$

If  $t_{\text{calc}} < t_{\text{tab}}$ , there is no statistical difference between  $\bar{x}$  and  $\mu^*$  and no systematic errors are observed at the specified confidence level. Equivalently, if  $\mu^*$  is encompassed in the confidence interval of  $\bar{x}$ , there is no statistical difference at the specified confidence level.

**(cont.)** A brass QC with a known copper content of  $(91.75 \pm 0.11)\%$  was analyzed and found to contain  $(92.2 \pm 0.5)\%$  copper, both at the 95% confidence level.

Ignoring the uncertainty in the QC,  $t_{\text{calc}}$  is determined to be 2.413.  $t_{\text{calc}}$  is lower than  $t_{\text{tab}}$ , 2.776, so there is no statistical difference at the 95% confidence level. No systematic errors were observed at the specified confidence level. Equivalently, the known QC value is encompassed within the confidence interval of the QC: 91.7% to 92.7%. Again, there is no statistical difference.

A statistical difference is found at the 90% confidence level.

Calculations involving uncertainty in both the experimental and known value is discussed in ***Comparing multiple data sets***, below.

Percent error is another common, but not statistical, calculation that measures deviation from the true value. Unlike the  $t$ -test, percent error provides information regarding the direction of a systematic error.

$$\% \text{ error} = \left( \frac{\text{experimental} - \text{actual}}{\text{actual}} \right) 100\% = \left( \frac{\bar{x} - \mu}{\mu} \right) 100\%$$

### Sources of uncertainty (ANOVA)

The square of the standard deviation is the variance. Variance is additive for normal distributions, making it possible to determine the magnitude of various sources of uncertainty. This analysis is often called ANalysis Of VAriance (ANOVA).

$$V = s^2$$

Figure 2 shows how a sample can be analyzed to determine the contributions from sampling, preparation, and analysis to the total uncertainty. The variance in **A** is due to the analysis only; the variance in **B** is due to analysis and preparation; the variance in **C** is the total variance of all processes. Because of the additive nature of variance,

$$V_{\text{total}} = V_{\text{sampling}} + V_{\text{preparation}} + V_{\text{measurement}}$$

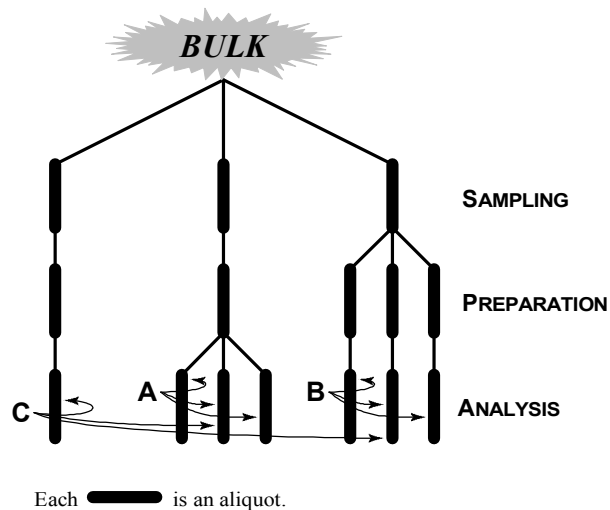


Figure 2. Sample flow-chart for an ANOVA analysis of a process.

### Propagation of uncertainty

Propagation of uncertainty is additive for variance but we more commonly work with standard deviation. The formula for propagating uncertainty through an arbitrary function,  $z = f(x, y, \dots)$ , is given by

$$s_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 s_y^2 + \dots + \left(\frac{\partial z}{\partial x}\right)\left(\frac{\partial z}{\partial y}\right) s_{xy} + \dots} \approx \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 s_y^2 + \dots}$$

The covariance,  $s_{ij}$ , is dependent on two variables and therefore more difficult to determine. All too often, covariance is not calculated (ignored). Note the likely confusion between  $s_i$ ,  $V_i$ , and  $s_{ij}$ :  $s_i$  is the standard deviation,  $V_i$  is the variance ( $V_i = s_i^2$ ), and  $s_{ij}$  is the covariance. Propagation of uncertainty functions for common mathematical operations are given below. Those where covariance is ignored have ' $\approx$ '.

Operation	Uncertainty	Operation	Uncertainty
$z = x + y$	$s_z^2 = s_x^2 + s_y^2$	$z = \ln(x)$	$s_z = \frac{s_x}{x}$
$z = x - y$		$z = \log(x)$	$s_z = \frac{1}{\ln(10)} \frac{s_x}{x}$
$z = x \cdot y$	$\left(\frac{s_z}{z}\right)^2 \approx \left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2$	$z = e^x$	$\frac{s_z}{z} = s_x$
$z = x/y$		$z = 10^x$	$\frac{s_z}{z} = \ln(10)s_x$
$z = x^a = x \cdot x \cdot \dots$	$\frac{s_z}{z} = a \frac{s_x}{x}$		<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <math>\uparrow</math> base e         </div> <div style="text-align: center;"> <math>\uparrow</math> base 10         </div> </div>

### Comparing multiple data sets

The  $F$ -test determines if the *variances* of two data sets ( $a$  and  $b$ ) are the same. Based on the results of the  $F$ -test, the  $t$ -test can be used to determine if the *means* of two data sets are the same. Alternatively, the  $F$ -test can be used to determine the significance of individual parameters in a model (non-linear curve fitting, for example.)

$$F_{\text{calc}} = \frac{V_a}{V_b} = \frac{s_a^2}{s_b^2} \quad a \text{ and } b \text{ are chosen so that } F_{\text{calc}} > 1.$$

$F_{\text{tab}}$  is then looked up for a specified confidence level. Unless there is a reason to believe otherwise, the two-tailed tabulated value is used. If  $F_{\text{calc}} < F_{\text{tab}}$ , we can say, "There is no statistical difference between the distributions at the specified confidence level.", and use a pooled standard deviation,  $s_{\text{pooled}}$ , in further calculations. Otherwise, individual standard deviations must be used and the degrees of freedom of  $t_{\text{tab}}$  must be calculated separately.

#### Passed F-test

$$s_{\text{pooled}} = \sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}}$$

$$t_{\text{calc}} = \frac{|\bar{x}_a - \bar{x}_b|}{s_{\text{pooled}} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}}$$

#### Failed F-test

$$t_{\text{calc}} = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

$$d.f. = \left[ \frac{\left( \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b} \right)^2}{\frac{\left( \frac{s_a^2}{n_a} \right)^2}{n_a + 1} + \frac{\left( \frac{s_b^2}{n_b} \right)^2}{n_b + 1}} \right] - 2$$

(cont.) To account for the uncertainty in the QC,  $F_{\text{calc}} = 0.417^2/0.056^2 = 55.4$ . (The quoted uncertainty is at the 95 % confidence level; assuming that an infinite number of analyses were conducted, the standard deviation was obtained by dividing by  $t(95\%, \infty) = 1.960$ .)  $F_{\text{calc}}$  is greater than  $F_{\text{tab}}$  at the 95 % confidence level ( $F_{\text{tab}(4, \infty)} = 2.786$ ); the two samples are not from the same population.

$t_{\text{calc}} = 2.413$ , which is lower than  $t_{\text{tab}}$ , 2.776, so there is no statistical difference at the 95 % confidence level.

NB: Both calculations for  $t_{\text{calc}}$  return the same value. If the number of replicates used to determine the QC is known,  $t_{\text{calc}}$  will differ.

*The examples on the next few pages illustrate how statistics is applied to more complicated systems. The examples can be omitted if desired. Beyond the examples are tabulated  $Q$ ,  $t$ , and  $F$  values. These tables and this complete document are also available from [www.consol.ca](http://www.consol.ca) (Teaching link).*

## Additional Examples

### Linear regression

The formulae required to conduct a linear regression analysis are beyond the scope of this summary. Furthermore, they are not needed: calculators, spreadsheets, and scientific software packages have these functions built-in. Not all programs determine the uncertainty associated with values read from a linear regression. I make an Excel spreadsheet available to assist with this calculation: [www.consol.ca](http://www.consol.ca) (Teaching link). The spreadsheet also contains the statistical tables for the tests discussed *supra*.

Linear regression analysis assumes that there is no uncertainty in the  $x$ -coordinate of each data point and that the uncertainty in the  $y$ -coordinate is constant for all data points. Figure 3 shows a typical linear calibration curve, the uncertainty in the regression, and the uncertainty in a single observation of a sample, all at the 95 % confidence level. The regression uncertainty is non-linear, being smaller near the center of the data set and larger towards the extrema. The uncertainty for a single observation of a sample is the same as the regression uncertainty; It is less for multiple measurements on the sample. The coefficient of determination,  $R^2$ , is a measure of the 'goodness of fit' of the fitting function to the experimental data points.

Non-linear mathematical transformations also skew the uncertainty distribution. To obtain statistically significant results, the data must be weighted with the relative uncertainty or the uncertainty determined from a non-linear calibration curve of the original data. Uncertainty analysis from non-linear calibration curves is beyond the scope of this summary.<sup>5</sup>

### Detection limits

The detection limit (DL) is defined as "the minimum single result [that], with a stated probability, can be distinguished from a suitable blank value."<sup>6</sup> In other words, the signal cannot be within the confidence interval of the blank, or a  $t$ -test between the blank and the sample must fail. In this case, a one-tailed  $t$ -value is used since the blank theoretically represents a minimum signal.

$$y_{DL} = y_{blank} + t s_b$$

In Figure 3, the minimum signal distinguishable from the blank is 0.072, which corresponds to a detection limit of 4.2  $\mu\text{mol/L}$ .

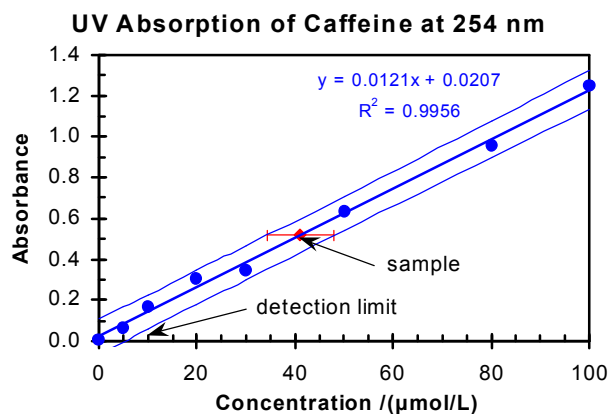


Figure 3. Caffeine calibration curve (solid line), linear regression uncertainty (light line), and single observation of a sample. All uncertainties at the 95 % confidence level.

## Simplified uncertainty for linear regression analyses

*I include this section with mixed feelings. Mostly, it is included to correct a commonly taught method that is grossly inaccurate.*

The formulae for statistically determining the uncertainty of a data point read from a linear regression analysis are complicated. Often, an estimate of the uncertainty is satisfactory and ideally, should be calculable from the uncertainty in the slope and intercept, which are provided by most programs.

**A POOR METHOD** (which is all too commonly used) involves plotting the minimum and maximum lines from

$$y_{\min}^{\max} = (m \pm s_m)x + (b \pm s_b) \text{ or}$$

$$y_{\min}^{\max} = (m \pm s_m)x + (b \mp s_b)$$

Although seemingly intuitive, the results differs significantly from the correct uncertainty as shown in Figure 4a. Contrary to the correct uncertainty, there exists a point where the uncertainty is supposedly zero and the uncertainty increases linearly in both directions from this point. Changing the  $\pm$  to  $\mp$  only changes the location of the 'zero-uncertainty' point.

**A BETTER METHOD** is obtained by simplifying the correct uncertainty formula. The resulting formula is a vertical shift of the best fit line and is in good agreement with the correct uncertainty formula (Figure 4b). The equations for the minimum and maximum uncertainty are

$$y_{\min}^{\max} = mx + (b \pm 1.5s_b)$$

The factor of 1.5 scales the uncertainty to minimize the errors of the approximations.<sup>7</sup> The resulting unknown uncertainty is calculable from

$$s_x = \frac{1.5s_b}{m}$$

$t$  can be included,  $t(1.5s_b)$ , to determine the uncertainty at a desired confidence level.

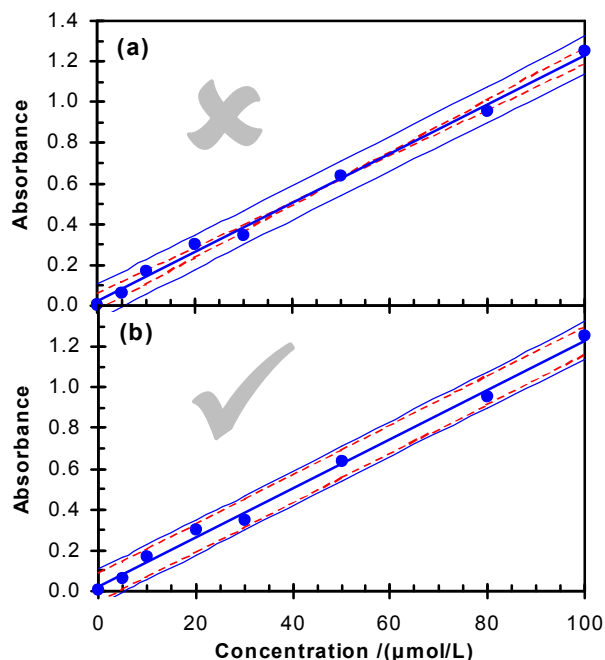


Figure 4. Simplified methods of estimating the linear regression uncertainty. Best fit (solid line), correct uncertainty (light line), and estimated uncertainty (dashed line). All uncertainties at the 95 % confidence level.

## Spectral data

We are often interested in the peaks in spectra: spectroscopy, mass spectrometry, chromatography, nmr, polarography, etc. These peaks often have a gaussian or near-gaussian (lorentzian, voigt) profiles. A convenient measure of the uncertainty is the full-width at half-maximum (FWHM). For a gaussian distribution,  $\text{FWHM} = 2\sqrt{2\ln(2)}\sigma \approx 2.36\sigma$ . Half this value,  $1.18\sigma$ , is the half-width at half-maximum (HWHM). Reporting the uncertainty as  $\pm\text{FWHM}$  corresponds to, at most, the 76 % confidence level ( $t(76\%, \infty) \approx 1.18$ ). ( $\pm\text{FWHM}$  corresponds to the 98 % confidence level.) Most peaks are comprised of fewer than an infinite number of observations, which is considered in the example below. By determining the standard deviation from the FWHM, the uncertainty at any confidence level can be determined. Alternatively, the centroid and standard deviation ( $x_0, s$ ) can be determined explicitly by fitting a gaussian function to the data

**Example:** High-resolution spectroscopy results in many peaks. Figure 5 shows a peak that is comprised of seven statistically significant data points (i.e., above the statistical uncertainty of the baseline). The FWHM is estimated as 0.008 nm and, assuming a gaussian like distribution, the standard deviation determined.  $t(98\%, 5) \approx 3.365$

$$\pm \text{FWHM} = 0.008 \text{ nm} = 3.365 s$$

$$s = 0.0024 \text{ nm}$$

Calculation of the centroid and standard deviation ( $x_0, s$ ) use two degrees of freedom. The uncertainty of the centroid at the 95 % confidence level is found to be

$$\begin{aligned} \Delta x_0 &= t s_{x_0} = \frac{t s}{\sqrt{n}} \\ &= \frac{2.571 \cdot 0.0024 \text{ nm}}{\sqrt{7}} = 0.0020 \text{ nm} \end{aligned}$$

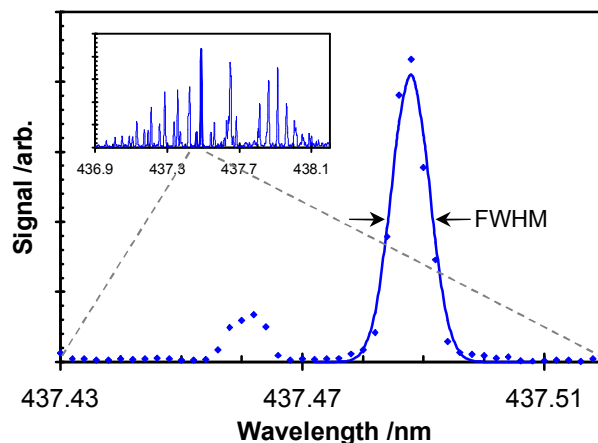


Figure 5. Gaussian distribution (solid line) fit to a single rovibronic transition in the  $C_2$  transition (points) near 437.6 nm. (from our laboratory)

### General reading

- Harris D.C. *Quantitative Chemical Analysis*, 6th ed., W. H. Freeman and Company, New York, **2002**.  
Harvey, D. *Modern Analytical Chemistry*, McGraw-Hill, New York, **2000**.

### References

1. Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry*, 6th ed., McGraw-Hill, New York, **1996**.
2. Taylor, B., *NIST Special Publication 811: Guidelines for the use of the International System of Units*, National Institute of Standards and Technology, **1995**.
3. *NIST Standard Reference Materials*, [www.nist.gov/srm/](http://www.nist.gov/srm/), January **2002**.
4. The Central Limit Theorem states that the distribution resulting from an infinite number of small independent influences (of any form) on a system will be gaussian.
5. See, for example, de Levie, R., *J. Chem. Ed.*, **1999**, 76, 1594.
6. International Union of Pure and Applied Chemistry (IUPAC) Goldbook, [www.iupac.org/publications/compendium/index.html](http://www.iupac.org/publications/compendium/index.html), March **2002**.
7. Jensen, R.H., unpublished results.

# Statistical Tables

**Table 1.** Tabulated values for the Q-test.

<i>n</i>	68%	90%	95%	98%	99%
3	0.822	0.941	0.970	0.988	0.994
4	0.603	0.765	0.829	0.889	0.926
5	0.488	0.642	0.710	0.780	0.821
6	0.421	0.560	0.625	0.698	0.740
7	0.375	0.507	0.568	0.637	0.680
8	0.343	0.468	0.526	0.590	0.634
9	0.319	0.437	0.493	0.555	0.598
10	0.299	0.412	0.466	0.527	0.568
12	0.271	0.375	0.425	0.480	0.518
14	0.250	0.350	0.397	0.447	0.483
16	0.234	0.329	0.376	0.422	0.460
18	0.223	0.314	0.358	0.408	0.438
20	0.213	0.300	0.343	0.392	0.420

**Table 2.** Tabulated values for the one and two-tailed *t*-tests.

D.F.	One-Tailed <i>t</i> -Test					Two-Tailed <i>t</i> -Test				
	68%	90%	95%	98%	99%	68%	90%	95%	98%	99%
1	0.635	3.078	6.314	15.894	31.821	1.819	6.314	12.706	31.821	63.656
2	0.546	1.886	2.920	4.849	6.965	1.312	2.920	4.303	6.965	9.925
3	0.518	1.638	2.353	3.482	4.541	1.189	2.353	3.182	4.541	5.841
4	0.505	1.533	2.132	2.999	3.747	1.134	2.132	2.776	3.747	4.604
5	0.497	1.476	2.015	2.757	3.365	1.104	2.015	2.571	3.365	4.032
6	0.492	1.440	1.943	2.612	3.143	1.084	1.943	2.447	3.143	3.707
7	0.489	1.415	1.895	2.517	2.998	1.070	1.895	2.365	2.998	3.499
8	0.486	1.397	1.860	2.449	2.896	1.060	1.860	2.306	2.896	3.355
9	0.484	1.383	1.833	2.398	2.821	1.053	1.833	2.262	2.821	3.250
10	0.482	1.372	1.812	2.359	2.764	1.046	1.812	2.228	2.764	3.169
12	0.480	1.356	1.782	2.303	2.681	1.037	1.782	2.179	2.681	3.055
14	0.478	1.345	1.761	2.264	2.624	1.031	1.761	2.145	2.624	2.977
16	0.477	1.337	1.746	2.235	2.583	1.026	1.746	2.120	2.583	2.921
18	0.476	1.330	1.734	2.214	2.552	1.023	1.734	2.101	2.552	2.878
20	0.475	1.325	1.725	2.197	2.528	1.020	1.725	2.086	2.528	2.845
25	0.473	1.316	1.708	2.167	2.485	1.015	1.708	2.060	2.485	2.787
30	0.472	1.310	1.697	2.147	2.457	1.011	1.697	2.042	2.457	2.750
40	0.471	1.303	1.684	2.123	2.423	1.007	1.684	2.021	2.423	2.704
50	0.471	1.299	1.676	2.109	2.403	1.004	1.676	2.009	2.403	2.678
75	0.470	1.293	1.665	2.090	2.377	1.001	1.665	1.992	2.377	2.643
100	0.469	1.290	1.660	2.081	2.364	0.999	1.660	1.984	2.364	2.626
200	0.468	1.286	1.653	2.067	2.345	0.997	1.653	1.972	2.345	2.601
500	0.468	1.283	1.648	2.059	2.334	0.995	1.648	1.965	2.334	2.586
∞	0.468	1.282	1.645	2.054	2.326	0.994	1.645	1.960	2.326	2.576

**Table 3.** Tabulated values for the two-tailed *F*-test.

<b>95%</b>		<b>Degrees of Freedom: Numerator</b>																							
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>16</b>	<b>18</b>	<b>20</b>	<b>25</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>75</b>	<b>100</b>	<b>200</b>	<b>500</b>	<b>∞</b>
<b>Degrees of Freedom: Denominator</b>	<b>1</b>	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	976.7	982.5	986.9	990.3	993.1	998.1	1001	1006	1008	1011	1013	1016	1017	1018
	<b>2</b>	38.51	39	39.17	39.25	39.3	39.33	39.36	39.37	39.39	39.4	39.41	39.43	39.44	39.44	39.45	39.46	39.46	39.47	39.48	39.48	39.49	39.49	39.5	39.5
	<b>3</b>	17.44	16.04	15.44	15.1	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.28	14.23	14.2	14.17	14.12	14.08	14.04	14.01	13.97	13.96	13.93	13.91	13.9
	<b>4</b>	12.22	10.65	9.979	9.604	9.364	9.197	9.074	8.98	8.905	8.844	8.751	8.684	8.633	8.592	8.56	8.501	8.461	8.411	8.381	8.34	8.319	8.288	8.27	8.257
	<b>5</b>	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.456	6.403	6.362	6.329	6.268	6.227	6.175	6.144	6.101	6.08	6.048	6.028	6.015
	<b>6</b>	8.813	7.26	6.599	6.227	5.988	5.82	5.695	5.6	5.523	5.461	5.366	5.297	5.244	5.202	5.168	5.107	5.065	5.012	4.98	4.937	4.915	4.882	4.862	4.849
	<b>7</b>	8.073	6.542	5.89	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.596	4.543	4.501	4.467	4.405	4.362	4.309	4.276	4.232	4.21	4.176	4.156	4.142
	<b>8</b>	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.2	4.13	4.076	4.034	3.999	3.937	3.894	3.84	3.807	3.762	3.739	3.705	3.684	3.67
	<b>9</b>	7.209	5.715	5.078	4.718	4.484	4.32	4.197	4.102	4.026	3.964	3.868	3.798	3.744	3.701	3.667	3.604	3.56	3.505	3.472	3.426	3.403	3.368	3.347	3.333
	<b>10</b>	6.937	5.456	4.826	4.468	4.236	4.072	3.95	3.855	3.779	3.717	3.621	3.55	3.496	3.453	3.419	3.355	3.311	3.255	3.221	3.175	3.152	3.116	3.094	3.08
	<b>12</b>	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.206	3.152	3.108	3.073	3.008	2.963	2.906	2.871	2.824	2.8	2.763	2.74	2.725
	<b>14</b>	6.298	4.857	4.242	3.892	3.663	3.501	3.38	3.285	3.209	3.147	3.05	2.979	2.923	2.879	2.844	2.778	2.732	2.674	2.638	2.59	2.565	2.526	2.503	2.487
	<b>16</b>	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.817	2.761	2.717	2.681	2.614	2.568	2.509	2.472	2.422	2.396	2.357	2.333	2.316
	<b>18</b>	5.978	4.56	3.954	3.608	3.382	3.221	3.1	3.005	2.929	2.866	2.769	2.696	2.64	2.596	2.559	2.491	2.445	2.384	2.347	2.296	2.269	2.229	2.204	2.187
	<b>20</b>	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.676	2.603	2.547	2.501	2.464	2.396	2.349	2.287	2.249	2.197	2.17	2.128	2.103	2.085
	<b>25</b>	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.515	2.441	2.384	2.338	2.3	2.23	2.182	2.118	2.079	2.024	1.996	1.952	1.924	1.906
	<b>30</b>	5.568	4.182	3.589	3.25	3.026	2.867	2.746	2.651	2.575	2.511	2.412	2.338	2.28	2.233	2.195	2.124	2.074	2.009	1.968	1.911	1.882	1.835	1.806	1.787
	<b>40</b>	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.288	2.213	2.154	2.107	2.068	1.994	1.943	1.875	1.832	1.772	1.741	1.691	1.659	1.637
	<b>50</b>	5.34	3.975	3.39	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.216	2.14	2.081	2.033	1.993	1.919	1.866	1.796	1.752	1.689	1.656	1.603	1.569	1.545
	<b>75</b>	5.232	3.876	3.296	2.962	2.741	2.582	2.461	2.366	2.289	2.224	2.123	2.046	1.986	1.937	1.896	1.819	1.765	1.692	1.645	1.578	1.542	1.483	1.444	1.417
<b>100</b>	5.179	3.828	3.25	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.077	2	1.939	1.89	1.849	1.77	1.715	1.64	1.592	1.522	1.483	1.42	1.378	1.347	
<b>200</b>	5.1	3.758	3.182	2.85	2.63	2.472	2.351	2.256	2.178	2.113	2.01	1.932	1.87	1.82	1.778	1.698	1.64	1.562	1.511	1.435	1.393	1.32	1.269	1.229	
<b>500</b>	5.054	3.716	3.142	2.811	2.592	2.434	2.313	2.217	2.139	2.074	1.971	1.892	1.83	1.779	1.736	1.655	1.596	1.515	1.462	1.381	1.336	1.254	1.192	1.137	
<b>∞</b>	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048	1.945	1.866	1.803	1.752	1.709	1.626	1.566	1.484	1.429	1.345	1.296	1.206	1.128	1	

# Units of Measure and Conversions

Quantity	SI Unit		Often used as
	Name	Abbreviation	
mass	kilogram	kg	kg, g, mg, $\mu$ g, ng
length	meter	m	km, m, cm, mm, nm
time	second	s	s, ms, $\mu$ s
temperature	Kelvin	K	K, mK
electric current	Amp	A	A, mA, $\mu$ A
luminosity	candela	cd	cd, mcd, $\mu$ cd
amount of substance	mole	mol	kmol, mol, mmol, $\mu$ mol

SI Prefix	Symbol	Value		Common Name
		Exponential	Full	
peta	P	$10^{15}$	1 000 000 000 000 000	
tera	T	$10^{12}$	1 000 000 000 000	trillion
giga	G	$10^9$	1 000 000 000	billion
mega	M	$10^6$	1 000 000	million
kilo	k	$10^3$	1 000	thousand
—		1	1	
deci	d	$10^{-1}$	0.1	
centi	c	$10^{-2}$	0.01	parts per hundred; %
milli	m	$10^{-3}$	0.001	parts per thousand (ppt)
micro	$\mu$	$10^{-6}$	0.000 001	parts per million (ppm)
nano	n	$10^{-9}$	0.000 000 001	parts per billion (ppb)
pico	p	$10^{-12}$	0.000 000 000 001	parts per trillion (ppt; pptr)
femto	f	$10^{-15}$	0.000 000 000 000 001	

Because the density of water is approximately 1 g/mL and the majority of analytical chemistry deals with aqueous solutions, the following approximate equivalents are found. They are exact when the density is exactly 1.000 g/mL.

**ppm**  $\approx$  mg/L;  $\mu$ g/mL; mg/kg;  $\mu$ g/g

**ppb**  $\approx$   $\mu$ g/L; ng/mL;  $\mu$ g/kg; ng/g

**pptr**  $\approx$  ng/L; pg/mL; ng/kg; pg/g