Tutorial Review

Calculating Standard Deviations and Confidence Intervals with a Universally Applicable Spreadsheet Technique

J. Kragten

Laboratory of Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

A quick and universally applicable spreadsheet method is outlined for the calculation of standard deviations based on the general formula for error propagation:

$$s_R^2 = \left(\frac{\partial R}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial R}{\partial y}\right)^2 s_y^2 + \left(\frac{\partial R}{\partial z}\right)^2 s_z^2 + \ldots$$

With this method, standard deviations are calculated numerically without violating the condition of mutual independence, with a substantial time gain and with no risk of calculating errors. Satterthwaite's approximation of the degrees of freedom is a logical extension of the technique with which confidence intervals can be easily established. Direct insight is obtained about the separate contributions of the different error sources.

Keywords: Numerical method; error propagation; estimated degrees of freedom; standard deviation; confidence interval

Introduction

Quality improvement of laboratory results forces us to look critically, more than ever before, at our statistical manipulations. The availability of computers avoids time-consuming calculations and permits calculations free from errors. However, there is still no guarantee that the final results are correct. For instance the estimation of standard deviations in error propagation is a field where 'rules' are often applied wrongly. It is well known that assumptions have been made in the derivation of these rules and that the underlying conditions have to be satisfied. Too often, however, the conditions are violated unintentionally and mistakes are made. This can be explained as follows. The simple rules

$$s_R^2 = s_x^2 + s_y^2 \quad \text{when} \quad R = x \pm y$$
$$\left(\frac{s_R}{R}\right)^2 = \left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 \quad \text{when} \quad \left.\begin{array}{c} R = xy \\ R = x/y \end{array}\right\} \quad (1)$$

are well known. They hold only for mutually independent variables. In practice, however, many observations are mutually independent, which makes most people think that the condition of mutual independence is satisfied automatically when the rules are applied, but this is not necessarily true. [For the quantities x, y, *etc.*, the best estimates (the mean values) will generally be substituted in the equation for R. It depends on the inquiry whether for s_x, s_y, *etc.*, the values of the population or of the mean ($s_{\bar{x}}$, $s_{\bar{y}}$, . . .) will be used in eqns. (1) and (2).]

The rules in eqn. (1) may only be applied when we are dealing with either an exclusive mix of additions and subtractions or with a combination of just divisions and multiplications. As soon as R is a more complicated function

of x, y, . . ., the simple rules lead to erroneous results. This will be shown with the calculation of the surface of a block: $R = 2(lb + bh + hl)$. Most workers will split R into the parts lb, bh and hl. The rules are applied to these separate parts and the standard deviations of these separate parts are obtained. Eventually the separate parts are summed to obtain R and the simple error propagation rules are applied again to find s_R. At this point the error is made: commonly the separate parts of R have some variables in common and hence are mutually dependent. (Use of the word *correlation* is restricted to covariance between measured quantities. Terms containing the same variable in a mathematical relationship will be called *dependent*.) The block-surface $R = 2(lb + bh + hl)$ is a good example with the product terms lb, bh and hl sharing b, h and l. Application of the simple rules generates a standard deviation s_R which is $\sqrt{2}$ too low. In more complicated cases this factor may even be as large as 10.

The only way to get rid of these errors is by using the general eqn. (2), in which R is the calculated Result, and x, y, z, . . . are the measured values whose standard deviations and covariances (usually of the mean) are known from the experiments.

$$s_R^2 = \left(\frac{\partial R}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial R}{\partial y}\right)^2 s_y^2 + \left(\frac{\partial R}{\partial z}\right)^2 s_z^2 + \ldots$$
$$+ 2\rho_{xy}\left(\frac{\partial R}{\partial x}\right)\left(\frac{\partial R}{\partial y}\right)s_x s_y + \ldots \quad (2)$$

The correlation coefficients ρ_{xy}, *etc.*, are defined as

$$\rho_{xy} = \frac{\text{covariance}\,(x, y)}{\sqrt{\text{var}\,(x)\,\text{var}\,(y)}}$$

In practice, most cross-product terms vanish because $\rho \approx 0$.

There is an aversion to applying eqn. (2) because of the complexity of its calculations. When eqn. (2) is used in the usual analytical way, we deal with n independent measured values, n differentiations, $n + 1$ equations and $n^2 + n$ substitutions. All these manipulations are time consuming, sensitive to calculation errors and cannot be generalized and automated as they are connected with the specific function $R(x, y, z)$. This can all be prevented by making use of numerically operating spreadsheet programs.

In the following, a standardized scheme of manipulations is presented. The procedure is universally applicable, takes into account dependences within the equations, gives much less risk of calculation errors and is executed in a fraction of the time. Although the method can be performed fruitfully on a sheet of paper with the aid of a programmable pocket calculator, it is preferably done with spreadsheet programs such as LOTUS 123, QUATTRO and EXCEL to lower the risk of personal errors.

Spreadsheet Method

The calculation scheme is presented schematically in Table 1 for the specific case of four measured quantities x, y, z and u. It is not a restriction. When we deal with n quantities the number of columns is extended to $n + 1$ and the number of rows is extended correspondingly.

First the manipulations are outlined; later the logic will be explained. The procedure is as follows:

(a) Arrange the n measured quantitites $x, y, z, u, \ldots$, in the left-most column from the top down as a column vector.

(b) Calculate the value of R and put the result underneath the vector $x, y, z, u \ldots$. When using a spreadsheet program, the equation for R is entered, but the value of R is displayed.

(c) Enter the value of R into the cell underneath by referring to the previous cell and making the cell address absolute (or by transferring its value by hand).

(d) Put the equation of the difference of the last two cells in the following cell in the column (first the difference ΔR will be zero in the first column).

(e) Calculate the squared difference in the next cell, entering the equation referring to the cell above (= zero again).

(f) The next step is to copy all filled cells of the first column n times into the columns to the right ($n = 4$ here as we adopted four variables in this specific case). An $n \times n$ square matrix of measured values appears (inside the double-lined area). The calculated R, the initial value of R, the difference and its square appear underneath again.

So far all columns are still identical, but this is modified during the following steps.

(g) Put in the row above the double-lined matrix the standard deviations in column sequence (s_x, s_y, s_z, s_u, . . .).

(h) Add the standard deviations to the diagonal elements of the matrix as shown in Table 1.

In the copying step, equations have been copied as equations and values as values. When the standard deviations are added to the diagonal elements in the double-lined matrix, new values of R will appear in the first row under the matrix. One row lower all cells contain the (constant) initial value of R.

In the next row the differences of the last two cells are displayed. The change of R can be estimated with a MacLaurin series development. Generally, the changes of R are small, which implies that the higher order terms in the series development may be neglected. As only one variable is changed in a column of the matrix, only the first partial derivatives will appear in each cell. The sequential changes are equal to: $(\partial R/\partial x)s_x$, $(\partial R/\partial y)s_y$, $(\partial R/\partial z)s_z$ and $(\partial R/\partial u)s_u$. By taking the square of the differences we obtain the terms of eqn. (2). Summing all terms gives the variance of R required. When the measured values are correlated with ρ known, the row with squares can be extended with the corresponding cross-product terms. In the rest of this paper covariance will not be considered any longer. It is not essential for the message.

Discussion

The schematic calculation offers a number of advantages, as follows.

The spreadsheet method is universally applicable. Going from one calculation to another only the equation for R in the corresponding cell has to be adapted. The rest of the

---

Table 1 Schematic representation of the spreadsheet method

Calculation of basic $R(x, y, z, \ldots)$	Numerical differentiation by change of the diagonal elements with the absolute standard deviations			
	s_x	s_y	s_z	s_u
x	$x + s_x$	x	x	x
y	y	$y + s_y$	y	y
z	z	z	$z + s_z$	z
u	u	u	u	$u + s_u$
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
R (eqn.)	$R + \left(\dfrac{\partial R}{\partial x}\right)s_x$	$R + \left(\dfrac{\partial R}{\partial y}\right)s_y$	$R + \left(\dfrac{\partial R}{\partial z}\right)s_z$	$R + \left(\dfrac{\partial R}{\partial u}\right)s_u$
R (value)	R	R	R	R
$-$	$-$	$-$	$-$	$-$
$\Delta R = 0$	$\left(\dfrac{\partial R}{\partial x}\right)s_x$	$\left(\dfrac{\partial R}{\partial y}\right)s_y$	$\left(\dfrac{\partial R}{\partial z}\right)s_z$	$\left(\dfrac{\partial R}{\partial u}\right)s_u$
$s_R{}^2 =$	$\left\{\left(\dfrac{\partial R}{\partial x}\right)^2 s_x{}^2 \; + \right.$	$\left(\dfrac{\partial R}{\partial y}\right)^2 s_y{}^2 \; +$	$\left(\dfrac{\partial R}{\partial z}\right)^2 s_z{}^2 \; +$	$\left.\left(\dfrac{\partial R}{\partial u}\right)^2 s_u{}^2\right\}$

$$\frac{1}{v_R} = \frac{1}{v_x}\left[\frac{\left(\dfrac{\partial R}{\partial x}\right)^2 s_x{}^2}{s_R{}^2}\right]^2 + \frac{1}{v_y}\left[\frac{\left(\dfrac{\partial R}{\partial y}\right)^2 s_y{}^2}{s_R{}^2}\right]^2 + \frac{1}{v_z}\left[\frac{\left(\dfrac{\partial R}{\partial z}\right)^2 s_z{}^2}{s_R{}^2}\right]^2 + \frac{1}{v_u}\left[\frac{\left(\dfrac{\partial R}{\partial u}\right)^2 s_u{}^2}{s_R{}^2}\right]^2$$

Note: establishing the confidence interval of the true value of R, the best known values of $x, y, \ldots$ (= their mean values) are taken in the first column together with the standard deviations of the mean in the top row. Generally the inquiries of the investigation determine whether standard deviations of the mean or of the individual are substituted in the top row. The bars (in $\bar{x}$, $s_{\bar{x}}$, ...) have been left out for ease of readability.

manipulations, repeating R in value, taking the difference, squaring the differences and summing the squares, remain the same. The spreadsheet method is applicable in a standard manner to all error propagation calculations.

Direct insight is obtained about dominating contributions to s_R. Hence if the precision has to be improved, the related error source is immediately known. The influence of an improvement can simply be forecasted by changing s_x.

A low risk of calculation error exists. In the first column the equation for $R(x, y, z, u)$ is introduced. As soon as the displayed value of R is correct, the copying process will no longer introduce errors. There is no violation of the condition of mutual independence. The calculated standard deviation will be correct.

There is a substantial gain in time. The calculation of R always has to be performed. If its value has been found correctly in the corresponding cell, the rest of the calculation scheme can be performed in minutes as the only manipulations still to be performed are copying the column n times and adding the standard deviations to the elements on the diagonal. It can be noted that for the calculations a programmable pocket calculator and a piece of paper will serve as well, but it takes some more time in moderately complex calculations and has a larger risk of errors.

Intermediate results can be introduced in the row above R (equation) if they are entered in the cell as equations. (See example 1 in Table 2; the logic will then be clear.)

Different functions R_i related to the same measured $x, y, \ldots$ can be added from the the top down in the left-most column. Their s_{Ri} can be calculated in the same run (see example 2 in Table 3).

A linearity check is quickly performed. Eqn. (2) was derived with the assumption that R changes linearly within the ranges $s_x, s_y, \ldots$ (= higher order terms in the MacLaurin series are negligible). When eqn. (2) is applied in the usual way, linearity should be checked. In numerical calculations, however, the higher order terms do *not* vanish. Therefore, linearity is easily controlled by subtracting the s values from the diagonal elements instead of adding them. Note that changes of a few per cent in s_R are not important regarding the uncertainties that standard deviations usually have.

The degrees of freedom of s_R can be estimated simply in the last row of the spreadsheet by using the equation of Satterthwaite.[1,2,3] The equation (Table 1) holds for non-correlated variables and is elucidated in the next section.

Degrees of Freedom of S_R

Neglecting the correlation between the measured values, eqn. (2) changes into

$$s_R^2 = \left(\frac{\partial R}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial R}{\partial y}\right)^2 s_y^2 + \left(\frac{\partial R}{\partial z}\right)^2 s_z^2 + \ldots \quad (3)$$

Means of four values (or more) follow a normal distribution to a good approximation because of the limit theorem.[3] Consequently, their variances will follow a χ^2-distribution:

$$\chi_v^2 = v\frac{s^2}{\sigma^2} \Leftrightarrow s^2 = \sigma^2 \frac{\chi_v^2}{v} \quad (4)$$

where v is the number of degrees of freedom of s. If s is the standard deviation of the mean, $v = N\text{-}1$; in case of regression $v = N\text{-}k$ (k is the number of constants in the polynomial). The degrees of freedom of $s_x, s_y, \ldots$ are exactly known from the experiments. In contrast to s_x^2, etc., s_R^2 follows its own distribution, which deviates slightly from a χ^2 distribution. Satterthwaite[4] investigated the distribution of s_R^2 for a number of cases and concluded that for all cases of practical

interest the χ^2 distribution may be adopted for s_R^2. He found that when $v_x, v_y, v_z, \ldots$ are small, the approximation is worse, but when calculating the confidence interval of s_R the error always remains negligible. For larger v the error still decreases as the distributions of both s_x^2, etc., and s_R^2 approach a normal distribution (limit theorem). The conclusion is that all s^2 in eqn. (3) can be replaced by substituting eqn. (4):

$$\sigma_R^2\frac{\chi_R^2}{v_R} = \left(\frac{\partial R}{\partial x}\right)^2 \sigma_x^2\frac{\chi_x^2}{v_x} + \left(\frac{\partial R}{\partial y}\right)^2 \sigma_y^2\frac{\chi_y^2}{v_y}$$
$$+ \left(\frac{\partial R}{\partial z}\right)^2 \sigma_z^2\frac{\chi_z^2}{v_z} + \ldots \quad (5)$$

In eqn. (5) all χ^2 are statistically variant quantities; the other quantities are constants. The left-hand side of eqn. (5) may be considered as a calculation result, which depends on the statistical fluctuating quantities (the χ^2s) on the right-hand side. We can apply again the rules for error propagation and calculate the standard deviations on both sides of eqn. (5). As the standard deviation of the χ_v^2 itself is equal to $2v$, we obtain

$$\left(\frac{\sigma_R^2}{v_R}\right)^2 (2v_R) = \left[\left(\frac{\partial R}{\partial x}\right)^2\frac{\sigma_x^2}{v_x}\right]^2 (2v_x) + \left[\left(\frac{\partial R}{\partial y}\right)^2\frac{\sigma_y^2}{v_y}\right]^2 (2v_y)$$
$$+ \left[\left(\frac{\partial R}{\partial z}\right)^2\frac{\sigma_z^2}{v_z}\right]^2 (2v_z) + \ldots \quad (6)$$

If finally all σs are replaced by their estimators s, Satterthwaite's relationship for estimating v_R is obtained:

$$\frac{1}{v_R} = \frac{1}{v_x}\left[\frac{\left(\frac{\partial R}{\partial x}\right)^2 s_x^2}{s_R^2}\right]^2 + \frac{1}{v_y}\left[\frac{\left(\frac{\partial R}{\partial y}\right)^2 s_y^2}{s_R^2}\right]^2$$
$$+ \frac{1}{v_z}\left[\frac{\left(\frac{\partial R}{\partial z}\right)^2 s_z^2}{s_R^2}\right]^2 + \frac{1}{v_u}\left[\frac{\left(\frac{\partial R}{\partial u}\right)^2 s_u^2}{s_R^2}\right]^2 \quad (7)$$

This means that the reciprocal of v_R is equal to the weighted mean of the reciprocal vs in which the weighting factors are formed by the square of the relative contributions of all error sources to the variance of s_R. v_R is rounded off to the nearest lower integer.

The relative contributions are easily calculated in some extra rows at the bottom of the scheme from the contributions already present in the calculation scheme. Finally, v_R is used to find the limiting $t(95\%$, two-tailed) for establishing the confidence interval of R.

Some Examples From Practice

(1) A potentiometric measurement is performed seven times for the determination of Ag^+ in solution. During the experiment the room temperature changed from 22 to 23 °C. The value of E_0 is known from 41 previous experiments. The results are given in Table 2 together with their standard deviations.

The equations used are

$$[Ag^+] = 10^{\frac{(E - E_0)}{factor}}$$

$$factor = 58.1 + \frac{(t - 20)}{5}$$

From Table 2, it follows that the direct determination of $[Ag^+]$ from potential measurements is not precise. Improvement is hardly possible because the determination of E_0 is not precise from solution. If E is measured in order to follow

concentration changes, a corresponding scheme will show that the precision of changes is much better as the influence of E_0 is then eliminated. As a consequence, potentiometric titrations can be performed with high precision even if the potential jumps are not sharp.

(2) After a calibration procedure to establish the relationship between stimulus x and response y, the straight calibration line is used to establish in two different samples the unknown contents x_1 and x_2 from the corresponding responses y_1 and y_2. (Errors from drift, sample taking, sample pretreatment, etc., are assumed to be absent for simplification, but can easily be introduced.)

The line function is transformed into the orthogonal form $y = a + b(x - \bar{x}_z)$ in order to make the covariance between a and b zero. The variances of a and b are known from calibration. We use the reversed form $x(y)$ in the spreadsheet scheme:

$$x = \frac{(y - a)}{b} + \bar{x}_z \text{ (with } a = \bar{y}_z)$$

The experiment was undertaken to establish a possible difference between the two contents x_1 and x_2. Prospectively x_1 and x_2 will be compared routinely by applying eqn. (1). s_{x1} and s_{x2} are known (Table 3) and so the standard deviation $s_{\Delta x} = \sqrt{(s_{x1}^2 + s_{x2}^2)} = 0.040$. Compared with $\Delta x = 0.062$ from the significance test it then follows that there is no reason to reject the null hypothesis (H_0: $\mu_1 = \mu_2$).

However, this conclusion is wrong. It is overlooked that although y_1 and y_2 are not covariant, x_1 and x_2 are not mutually independent as they are found with the same calibration line. Hence application of the simple rules [eqn. (1)] is not permitted. It should be emphasized that all results following a specific calibration are related to the same calibration line and hence are covariant!

The correct answer is found by calculating both x_1, x_2 and $\Delta x = x_2 - x_1$ directly from the independent quantities a, b, y_1 and y_2 in one spreadsheet. The standard deviation of Δx now found is smaller, 0.026, and the correct conclusion from the significance test is that H_0 has to be rejected.

From the separate contributions of a, b, y_1 and y_2 to the standard deviation $s_{\Delta x}$, it will be obvious why $s_{\Delta x}$ is smaller: neither s_a nor s_b contributes substantially.

Looking at a graphical representation, it will be clear that x_1 and x_2 are dependent. Both x_1 and x_2 will change with a and b,

but a shift of the line does not change their difference Δx and a slight rotation will change Δx only marginally.

From this example, it is obvious that if all quantities R_i are properly related by their equations to the measured values at the top left, the spreadsheet method automatically takes into account the mathematical dependence between the variables.

(3) In ISO 6976-1984(E),[5] the calculation of the calorific value, density and relative density of natural gas is described. In the document numerous pages are dedicated to the analytical procedure for calculating the standard deviation of the calorific value s_w. Following this procedure, it is difficult to find the main sources of errors back in s_w. For this reason, an investigation was undertaken at Gasunie Research in which the numerical spreadsheet method was compared with the ISO 6976 procedure.[6] About 125 error sources were involved. Some were forecasted to be important, but turned out to be unimportant. Others were overlooked. The analytical approach took months for completion. The numerical method took a few days and gave much more insight. The error sources were grouped for convenience. The main source was the literature values of the physical constants. Second, the error in the determination of nitrogen was important, as the calorific value of this compound is zero. It is now proposed to introduce the spreadsheet technique in the ISO specifications.

Conclusions

The numerical spreadsheet technique has five main advantages. First, if all quantities to be calculated are properly related to the measured values, in the top of the left-most column, there will be no accidental violation of the condition of mutual independence during the calculation phase. Second, compared with the analytical method there is no risk of making errors in calculating and differentiating. If R has been found correctly in the left-most column, all other manipulations consist of copying columns and adding standard deviations to the diagonal elements with standard spreadsheet commands. Generally these manipulations are performed with a much smaller risk. Third, the standard manipulations of

Table 2 Potentiometric [Ag+] determination

		s_E 1.0 mV	s_{E_0} 1.0 mV	t 0.5 °C
$E =$	663.0 mV	664.0	663.0	663.0
$E_0 =$	779.5 mV	779.5	780.5	779.5
$t =$	22.5 °C	22.5	22.5	23.0
Factor =	58.6 mV	58.6 mV	58.6 mV	58.7 mV
[Ag+] =	0.010279	0.010691	0.009883	0.010359
[Ag+] =	0.010279	0.010279	0.010279	0.010279
$\Delta =$	0	0.000412	−0.000396	0.000080
$\Delta^2 =$	0	1.70×10^{-7}	1.57×10^{-7}	6.38×10^{-9}

$s_{Ag}^2 = 3.31 \times 10^{-7}$, hence $s_{Ag} = 5.78 \times 10^{-4}$ and [Ag+] $= 102.96 \times 10^{-4}$

$\nu =$	6 df	40 df	1 df
Weight =	0.260	0.222	0.004
Weight/ν =	0.0433	0.0055	0.004

$1/\nu_{Ag} = 0.0492$, $\nu_{Ag} = 20$, $t(95\%, 20 \text{ df}) = 2.09$

Confidence interval [Ag+] $= 0.0103 \pm 0.0012$ (95%, 20 df)

Table 3 Comparison of results after a single calibration

		s_a 0.051	s_b 0.014	s_{y_1} 0.045	s_{y_2} 0.045
a	12.63	12.681	12.63	12.63	12.63
b	2.43	2.43	2.444	2.43	2.43
y_1	13.10	13.10	13.10	13.145	13.10
y_2	13.25	13.25	13.25	13.25	13.295
x_1	5.193	5.173	5.192	5.212	5.193
x_1	5.193	5.193	5.193	5.193	5.193
Δ	0	−0.021	−0.001	0.019	0
Δ^2		4.3×10^{-4}	1.3×10^{-6}	$3.5^1 10^{-4}$	

$s_{x_1}^2 = 7.8 \times 10^{-4}$ $s_{x_1} = 0.028$

		s_a	s_b	s_{y_1}	s_{y_2}
x_2	5.255	5.234	5.254	5.255	5.274
x_2	5.255	5.255	5.255	5.255	5.255
Δ	0	−0.021	−0.001	0	0.019
Δ^2		4.3×10^{-4}	1.3×10^{-6}		3.5×10^{-4}

$s_{x_2}^2 = 7.8 \times 10^{-4}$ $s_{x_2} = 0.028$

		s_a	s_b	s_{y_1}	s_{y_2}
$x_2 - x_1$	0.062	0.06200	0.061	0.043	0.080
$x_2 - x_1$	0.062	0.06200	0.062	0.062	0.062
Δ	0	0.00000	$-3.6 \times^{-4}$	−0.019	0.019
Δ^2		0.00000	1.3×10^{-7}	3.5×10^{-4}	3.5×10^{-4}

$s_{\Delta x}^2 = 6.9 \times 10^{-4}$ $s_{\Delta r} = 0.026$

the technique are identical from case to case. Once familiar with these manipulations, the calculations can be performed in a fraction of the time required by the classical technique. Fourth, the spreadsheet table gives direct insight into the magnitude of each error contribution. Finally, estimating the degrees of freedom of the calculated result with the Satterthwaite's equation simply involves adding an additional row to the spreadsheet table.

References

1 Satterthwaite, F. E., *Biom. Bull.*, 1946, **2**, 110.
2 Burdick, R. K., and Graybill, F. A., *Confidence Intervals on Variance Components*, Marcel Dekker, New York, 1992, p. 30.
3 *Guide to the Expression of Uncertainty in Measurement*, ISO/ TAG4/WG3, 1st edn., ISO, Geneva, 1993.
4 Satterthwaite, F. E., *Psychometrika*, 1941, **6**, 309.
5 *Natural Gas—Calculation of Calorific Value, Density and Relative Density*, ISO 6976-1984(E), International Organization for Standardization, Geneva, 1984.
6 Kenter, R., Struis, M., and Smit, A. L. C., *Process Control Qual.*, 1991, **1**, 127.