Solution to measurement uncertainty challenge

Juris Meija¹

Published online: 14 September 2017
© Crown copyright in right of Canada 2017

The winner of the measurement uncertainty challenge (published in volume 409, issue 10) is Andrea Bazzano, Agenzia Regionale per la Protezione dell'Ambiente Ligure (ARPAL), Italy.

The award entitles the winner to select a Springer book of his choice up to a value of €100.

Our congratulations!

The analysis of sterling silver has produced information regarding the silver content in this alloy [1]. Although we are interested only in the mass fraction of silver in the alloy, let us first collect all that is known about this material:

1. The alloy is produced from pure silver and pure copper. Consequently, no other elements are expected in the final material.
2. The mass of each pin is 1 g.
3. Determination of silver from five pins gave the following results for the mass of silver in each pin: $m(\text{Ag}) = 0.844$ g, 0.888 g, 0.825 g, 0.907 g, 0.882 g, with standard deviation $s(m(\text{Ag})) = 0.034$ g.
4. Determination of copper from five pins gave the following results for the mass of copper in each pin: $m(\text{Cu}) = 0.060$ g, 0.096 g, 0.067 g, 0.075 g, 0.070 g, with standard deviation $s(m(\text{Cu})) = 0.014$ g.

It is clear that the available information is contradictory. For example, the average masses of silver and copper add up to 0.94 g, which is 6% below the stated mass of each pin. One path to reconcile the discrepancies is to inspect the data for potential outliers. The Dixon or Grubbs test for a single high outlier suggests that the copper measurements contain a possible outlier (0.096 g). Removal of this outlier, however, does not reconcile the discrepancy between the total mass of each pin and the sum of the silver and copper contents. In situations like these, the analyst is faced with making a decision on the likely reasons for such disagreements. It could be that the copper measurements are erroneous—after all, the results from five pins show 20% variability. It is also possible for the silver measurements to be in error, and we might doubt that each pin weighs 1 g or the fact that the alloy was indeed made from pure starting metals. On the contrary, one can interpret all of the information in a self-consistent manner by allowing that any observed discrepancies from the analysis of the ten pins is entirely due to inhomogeneity of the chemical composition between pins. Several ways to interpret the available data are summarized in Table 1, and they all lead to different measurement models and conclusions regarding the mass fraction of silver in the material.

Option A

This model provides rather a straightforward estimate of the mass fraction of silver: $w(\text{Ag}) = m(\text{Ag})/(1 \text{ g})$. The mean of the five observations is $w_A(\text{Ag}) = 0.869 \text{ g/g}$. The 95% confidence interval of $w_A(\text{Ag})$ can be obtained, assuming a normal distribution of the measurement errors. Since the actual (true)

This article is the solution to the Analytical Challenge to be found at <http://dx.doi.org/10.1007/s00216-017-0210-4>.

✉ Juris Meija
juris.meija@nrc-cnrc.gc.ca

¹ Measurement Science and Standards, National Research Council Canada, 1200 Montreal Road M-12, Ottawa, ON K1A 0R6, Canada

Table 1 Alternative approaches to data reduction depending on the available information used

Option	Source of information used				Measurement model
	$w_{\text{Ag}} + w_{\text{Cu}} = 1$	$m = 1 \text{ g}$	$m(\text{Ag})$	$m(\text{Cu})$	
A		×	×		$w_{\text{A}}(\text{Ag}) = m(\text{Ag})/(1 \text{ g})$
B		×		×	$w_{\text{B}}(\text{Ag}) = (1 - m(\text{Cu}))/(1 \text{ g})$
C	×		×		$w_{\text{C}}(\text{Ag}) = m(\text{Ag})/(m(\text{Ag}) + m(\text{Cu}))$
D	×	×	×	×	Homogeneity in $w(\text{Ag})$ and $w(\text{Cu})$ is assumed between the pins; $w_{\text{D}}(\text{Ag})$ is obtained from the fixed-effects statistical model; e.g., weighted average of $w_{\text{A}}(\text{Ag})$ and $w_{\text{B}}(\text{Ag})$
E	×	×	×	×	Inhomogeneity in $w(\text{Ag})$ and $w(\text{Cu})$ is assumed between the pins; $w_{\text{E}}(\text{Ag})$ and its uncertainty are obtained from bootstrap resampling of $m(\text{Ag})$ and $1 - m(\text{Cu})$ data

magnitude of the measurement errors is unknown, we use the quantiles of the t distribution with four degrees of freedom to obtain the expanded uncertainty of $w_{\text{A}}(\text{Ag})$:

$$U(w_{\text{A}}(\text{Ag})) = t_{p=1-\frac{0.05}{2}, v=4} \frac{s(m(\text{Ag}))}{\sqrt{5}} = 2.78 \frac{0.034}{\sqrt{5}} = 0.042 \text{ g/g}. \quad (1)$$

Alternatively, the doubt about the standard deviation $s(m(\text{Ag}))$ can be incorporated with use of the parametric bootstrap resampling method, where we draw samples from a normal distribution with a known mean (0.869 g/g) and unknown standard deviation. Here, the standard deviation itself is modeled as a random variable that follows a scaled and shifted chi-squared distribution to be consistent with the observed standard deviation:

$$s_i(m(\text{Ag})) = s(m(\text{Ag})) \sqrt{\frac{N-1}{\chi_{N-1}^2}}. \quad (2)$$

We proceed with random sampling (of $N = 5$ samples) from the normal distribution with the mean 0.869 g/g and standard uncertainty s_i as given in Eq. 2. The means of such samples provide 95% confidence interval $U(w_{\text{A}}(\text{Ag})) = 0.042 \text{ g/g}$, in agreement with the result shown in Eq. 1.

Option B

Calculations for this model follow the same route as in option A, and provide $w_{\text{B}}(\text{Ag}) = 0.926 \text{ g/g}$ with expanded uncertainty $U(w_{\text{B}}(\text{Ag})) = 0.017 \text{ g/g}$ with use of either Eq. 1 or Eq. 2.

Option C

To some, the most intuitive choice for data reduction of this dataset is to take into account both silver and copper determinations while disregarding the fact that the pins each weigh 1 g. The corresponding measurement model is

$$w_{\text{C}}(\text{Ag}) = \frac{m(\text{Ag})}{m(\text{Ag}) + m(\text{Cu})}. \quad (3)$$

Because the same variable appears in the equation, one can rewrite Eq. 3 to avoid the dependency problem between the numerator and the denominator:

$$w_{\text{C}}(\text{Ag}) = 1/(1 + m(\text{Cu})/m(\text{Ag})). \quad (4)$$

From Eq. 4, we obtain $w_{\text{C}}(\text{Ag}) = 0.922 \text{ g/g}$. Modeling the average results for both $m(\text{Ag})$ and $m(\text{Cu})$ as random variables with a t distribution, $m(\text{Ag}) \sim t(\mu = 0.869, s = 0.015, v = 4)$ and $m(\text{Cu}) \sim t(\mu = 0.074, s = 0.006, v = 4)$, we obtain expanded uncertainty $U(w_{\text{C}}(\text{Ag})) = 0.012 \text{ g/g}$. The parametric bootstrap method involving Eq. 2 to account for the uncertainty in the standard deviation of both copper and silver measurements provides an expanded uncertainty $U(w_{\text{C}}(\text{Ag})) = 0.017 \text{ g/g}$.

Peter mistakenly thought that both silver and copper measurements were made from the same five pins. He calculated the correlation between these data (0.578) and propagated the uncertainty by taking this into account (The NIST Uncertainty Machine at <http://uncertainty.nist.gov> is useful to do such calculations).

Option D

Option D combines all the available information while assuming the homogeneity of the pins. This can be achieved, for example, by our taking the average of the results w_{A} and w_{B} . For this purpose, we can choose the weighted average:

$$w_{\text{D}}(\text{Ag}) = \frac{u^{-2}(w_{\text{A}}) \cdot w_{\text{A}}(\text{Ag}) + u^{-2}(w_{\text{B}}) \cdot w_{\text{B}}(\text{Ag})}{u^{-2}(w_{\text{A}}) + u^{-2}(w_{\text{B}})}, \quad (5)$$

$$u^2(w_{\text{D}}(\text{Ag})) = \frac{1}{u^{-2}(w_{\text{A}}) + u^{-2}(w_{\text{B}})}. \quad (6)$$

Using the values $w_{\text{A}}(\text{Ag}) = 0.869(15) \text{ g/g}$ and $w_{\text{B}}(\text{Ag}) = 0.926(6) \text{ g/g}$, we obtain $w_{\text{D}}(\text{Ag}) = 0.918(6) \text{ g/g}$. The expanded

uncertainty of $w_D(\text{Ag})$ can be obtained by multiplication of the standard uncertainty and the corresponding t value. The Welch–Satterthwaite equation provides the approximate degree of freedom for the weighted average, $\nu = 5.2$, which corresponds to a 95% critical t value of 2.6 and expanded uncertainty $U(w_D(\text{Ag})) = 2.6 \times 0.006 = 0.016$ g/g. The weighted average is not without its shortcomings. Most notably, as explained earlier, the estimates of variance can be inaccurate when they are derived from a small number of replicates.

Option E

Option D assumes that all pins have identical mass fraction of silver, which might not be the case. A method that does not make such an assumption is nonparametric bootstrap resampling of the data. For this purpose, we can evaluate the uncertainty of the silver mass fraction from bootstrap resampling (with replacement) from the available set of ten results: five values of $m(\text{Ag})/(1 \text{ g})$ and five values of $(1 - m(\text{Cu}))/(1 \text{ g})$, which treats both silver and copper data on an equal footing. The average is calculated for each bootstrap sample and, after a sufficient number of such samples, the following summary statistics is obtained: $w_E(\text{Ag}) = 0.898$ g/g with expanded uncertainty $U(w_E(\text{Ag})) = 0.023$ g/g.

Summary

In summary, different choices on how to interpret the measurement results will lead to different uncertainty statements, [2] as shown in Table 2. The approaches outlined here are by no means the only choices that can be made, and many

Table 2 Alternative approaches to data reduction depending on the available information used

Option	Result, $w(\text{Ag})$	Expanded uncertainty, $U(w(\text{Ag}))$
A	0.869 g/g	0.042 g/g
B	0.926 g/g	0.017 g/g
C	0.922 g/g	0.017 g/g
D	0.918 g/g	0.016 g/g
E	0.898 g/g	0.023 g/g

readers might choose other ways to interpret this seemingly simple dataset. Indeed, as is noted in the *Guide to the Expression of Uncertainty in Measurement*, “the evaluation of measurement uncertainty is neither a routine task nor a purely mathematical one” [3]. Hence, measurement uncertainty is somewhat of a personal statement; it is *your* uncertainty, not *the* uncertainty.

References

1. Meija J. Measurement uncertainty challenge. *Anal Bioanal Chem.* 2017;409:2497.
2. Possolo A, Pintar A L. Plurality of Type A evaluations of uncertainty. *Metrologia* 2017;54:617.
3. Joint Committee for Guides in Metrology. Evaluation of measurement data – Guide to the expression of uncertainty in measurement, JCGM 100:2008. Joint Committee for Guides in Metrology; 2008