A worked example to estimate sampling precision & measurement uncertainty

Nearly all analysis requires the taking of a sample, a procedure which itself introduces uncertainty into the final test result. Hence a measurement uncertainty should cover both the uncertainties of sampling and analysis.

The following worked example quantifies the variance associated with sampling with particular reference to its precision due to random error.

Suppose that we need to take a number of soil samples from a contaminated site for the analysis of cadmium contents through a completely random sampling process and then analyse them in duplicate in also a completely random order. The results are summarized in Table 1.

As soil is often very heterogeneous in nature, we can safely expect the samples drawn are different. We want to know by how much they differ. Hence, we are not really interested in a significance test. The design of experiment is schematized in Figure 1.

Figure 1: An simple split experimental design for sampling and analytical variance

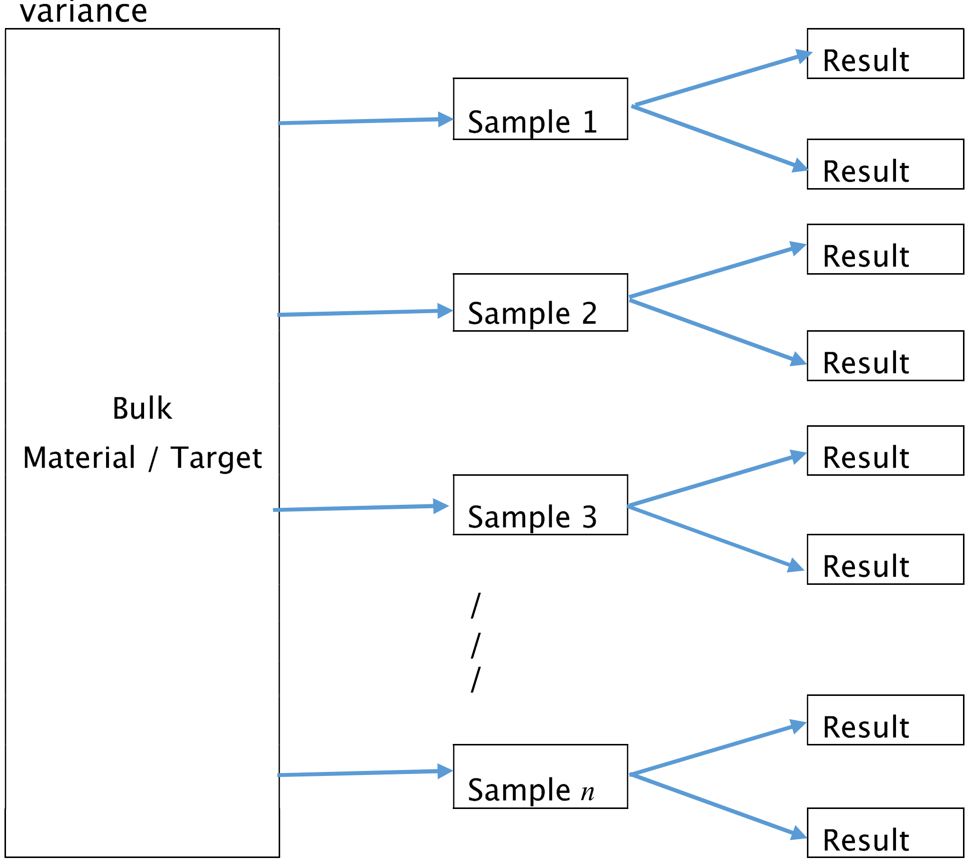

Table 1: The Cadmium results (in ppm) of 10 random samples taken from a field (ref: Michael Thompson & Philip Lowthian, page 58)

Soil sample	1	2	3	4	5	6	7	8	9	10
Result 1	11.8	6.4	11.9	12.2	7.5	6.4	10.1	11.3	14.0	16.5
Result 2	9.8	6.3	10.3	10.2	7.3	6.4	10.0	9.9	12.5	15.1
Mean =	10.8	6.35	11.1	11.2	7.4	6.4	10.05	10.6	13.25	15.8

One-way ANOVA technique can be applied in this case. We shall look into its within- and between-sample variations.

Estimation of Within sample (analysis) variance

In this case, we have n=10 number of samples with duplicate results (k = 2) for each sample. The grand mean of the data, $\bar{\bar{x}}$ is 10.295 ppm.

The deviation ($D_{i,1}$) of result 1 of sample i from the mean value of sample i is:

$$D_{i,1} = (x_{i,1} - \bar{x}_i)$$

Also, the deviation ($D_{i,2}$) of result 2 of sample i from the mean value of sample i is

$$D_{i,2} = (x_{i,2} - \bar{x}_i)$$

In fact, $D_{i,1} = D_{i,2}$. Therefore, the sum of squares of deviation (error) for sample i is:

$$SSE_i = (x_{i,1} - \bar{x}_i)^2 + (x_{i,2} - \bar{x}_i)^2 = 2(x_{i,1} - \bar{x}_i)^2$$

Hence, the total sum of squares of analysis error (within-sample) for all the 10 samples are:

$$SS_w(analysis) = \sum_{i=1}^{10} SSE_i$$

In this worked example, $SS_w(analysis)$ = 8.395, whilst the degree of freedom for within sample (between-analysis) = n x k – n = 10 x 2 – 10 = 10.

Therefore the mean square of analysis error $MS_w(analysis)$ = 8.395/10 = 0.8395.

Estimation of between-sample variance

To estimate the variation between samples, we study the deviation square $D^2$of the mean value of each sample from the overall or grand mean value, $\bar{\bar{x}}$.

Sample, i	Sample Mean $i,1$	D(Mean) Squares
1	10.8	0.255
2	6.35	15.563
3	11.1	0.648
4	11.2	0.819
5	7.4	8.381
6	6.4	15.171
7	10.05	0.060
8	10.6	0.093
9	13.25	8.732
10	15.8	30.305
Grand Mean $\bar{\bar{x}}$	10.30	

The sum of deviation squares of the above data = 80.0273 whilst the replicate k =2. Hence, the sum of squares (Between-samples) is:

$$SS_b = \sum_{i=1}^{10} k_i(\bar{x}_i - \bar{\bar{x}})^2 = 160.055$$

As there are 10 samples under consideration, the degree of freedom between-samples therefore is (10-1) or 9.

The mean square of between-samples $MS_b = \frac{SS_b}{df_b}$=17.784

Significance F-test evaluation

$F = \frac{MS_b}{MS_w}$ =21.18 which is larger than the critical value $F_{\alpha=0.05,v1=9,v2=10}$ of 3.02, indicating the samples under study were significantly different.

Indeed, the above values of MS_w and MS_b are confirmed by the MS Excel® Data Analysis tool as shown below:

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Soil 1	2	21.6	10.8	2
Soil 2	2	12.7	6.35	0.005
Soil 3	2	22.2	11.1	1.28
Soil 4	2	22.4	11.2	2
Soil 5	2	14.8	7.4	0.02
Soil 6	2	12.8	6.4	0
Soil 7	2	20.1	10.05	0.005
Soil 8	2	21.2	10.6	0.98
Soil 9	2	26.5	13.25	1.125
Soil 10	2	31.6	15.8	0.98

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Samples	160.0545	9	17.78383	21.18384	2.26552E-05	3.020383
Within samples (between analysis)	8.395	10	0.8395			
Total	168.4495	19				

Estimation of measurement uncertainty

From the above mean squares, we calculate the analytical standard deviation s_a as:

$$s_a = \sqrt{MS_w} = \sqrt{0.839} = 0.92 \text{ ppm}$$

The sampling standard deviation s_s is given by:

$$s_s = \sqrt{\frac{MS_b - MS_w}{k}} = \sqrt{\frac{17.78 - 0.84}{2}} = 2.91 \text{ ppm}$$

(Note that $k = 2$ because we did duplicate analysis.)

The total standard deviation for a combined operation of sampling (random) and analysis is going to be:

$$s_{Total} = \sqrt{s_s^2 + s_a^2} = \sqrt{2.91^2 + 0.92^2} = 3.05 \text{ ppm}$$

Remarks

The uncertainty of sampling bias, if any, was not considered in this case.

In fact, sampling bias is a contentious subject with many supporters and opponents. Some claim that sampling bias does not exist as the sampling protocols must have been agreed upon between the sampler and the customer. This implies that the sampling method is akin to an empirical analytical method, tailored to the fact that the sampling method defines the measurand or the analyte of interest.

But, in reality, sampling bias could arise like in this worked example. As the site was contaminated, the sampling tools might contaminate the samples taken if no proper precautions were taken to ensure their cleanliness. The sampler at site might possibly misinterpreted the sampling protocols and made systematic error in the sampling process.

References

- Michael Thompson & Philip Lowthian, *Notes on Statistics and Data Quality for Analytical Chemists* (Imperial College Press, 2011)

- Eurachem/CITAC Guide *Measurement Uncertainty arising from Sampling* (1st Edition 2007)