Estimating bias between two sampling methods

Measurement uncertainty has two main contributors, namely sampling uncertainty and analytical uncertainty, but most laboratory analysts tend to equate analytical uncertainty as its measurement uncertainty based on the sample received. This may be true when the target (population) lot sampled is homogeneous where every part of the target have an equal chance of being incorporated in the sample.

In practice, this is not always the case. Indeed, most sampling errors, except the preparation errors, are due to the material heterogeneity. There are two classes of material heterogeneity:

- Constituent heterogeneity – all natural materials are heterogeneous, i.e., they consist of different types of forms (metallic, molecular, ionic, grains, etc.).
- Distribution heterogeneity – if the material particles are not randomly distributed in the sampling target (lot, population) under study.

Therefore there is a potential uncertainty in the sampling process prior to the measurement process in the laboratory to be considered seriously.

The experimental design of sampling discussed in the recent blog, (*https://consultglp.com/2018/02/10/sampling-designs-for-uncertainty-estimation/*) was aimed to achieve the best sampling precision or minimize sampling error. ANOVA techniques are being used to study the between- and within-sampling and also between-analysis variances.

Like the evaluation of analytical uncertainty, we must address both uncertainty contributions, namely sampling precision and sampling bias, if any.

It is quite difficult to handle sampling bias adequately. It is also well known that sampling may be biased, for example

- by differential removal of materials,
- unintentional cross-contamination of sampling tools,
- misinterpretation of sampling protocols,
- inappropriate timing of sampling where temporal fluctuations occur, or
- by access restrictions.

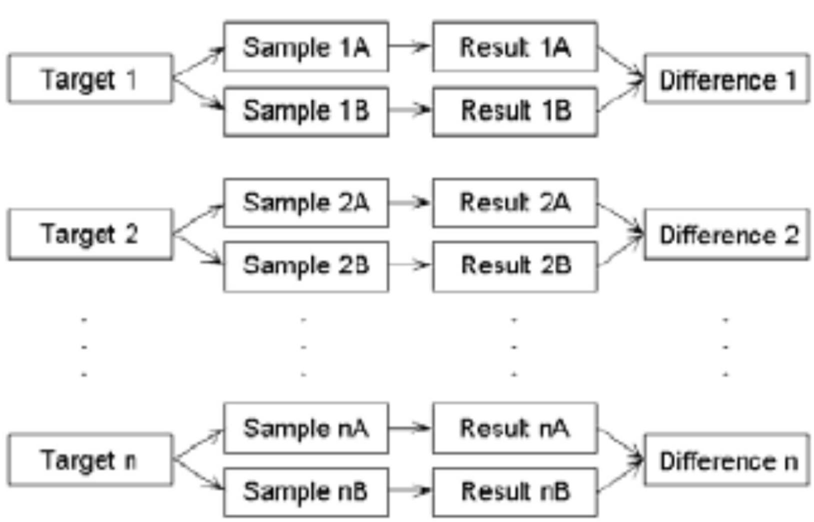
We can consider analogues of methods used to study bias in analytical methods as potential tools for handling sampling bias:

- In principle, we can use a certified sampling reference target (analogue of a CRM in analytical methods*)* or an established sampling method as a reference for method comparison*;*
- Inter-sampler studies with a single sampling protocol (analogue of a collaborative trial) to address between-sampler variation. This "reproducibility sampling variance" could be used as an extra term in the combined uncertainty. *However, this exercise can be costly to organize for field sampling.*

Eurachem/CITAC Guide "Measurement uncertainty arising from sampling" proposes a method to estimate bias between two sampling methods, by using *paired samples.* The sampling plan for such study is shown in Figure 1.

This is a design of experiment for bias estimation between two sampling methods (A and B) for a number of targets (preferably $n > 20$) by a single sampler, with single analysis carried out on the analysis of each randomly drawn sample. In this manner, any systematic error in the analysis is cancelled off. Significance test using the Student's t-test for paired comparison for the sampling is then carried out after the measurement results are obtained.

Figure 1: An experimental design to take samples by two sampling methods, one of which is a reference or standard method.



In this case, the null hypothesis H_o is that there are no differences in measurements by these two methods (i.e., Difference between the paired test results, d = 0) whilst the alternate hypothesis H_1 is that d>0. Use the equation (1) for the calculation of t– value and comparing it with the t–critical value with a degree of freedom n-1:

$$t = \frac{|\bar{x}_d|\sqrt{n}}{s_d} \qquad (1)$$

where $\bar{x}_d$ is the mean of difference of paired results,

s_d, the difference standard deviation and

n, the number of paired differences.

If the calculated t value is larger than the $t_{\alpha=0.05,v=n-1}$ critical value, Ho is rejected, inferring that there is a significant difference in performance between these two sampling methods by the same sampler.

Worked Example

A sampling method comparison exercise was conducted by randomly sampled 20 drums of a consignment of 300 drums (250-kg each) of a pesticide emulsifiable concentrate (EC) formulation by two sampling protocols (A and B), one of which was an approved method (Method B). Each sample drawn was analyzed once for its active ingredient, expressed as %w/w.

The raw data and the calculations of their squared differences of paired results for all targets were summarized in Table 1.

From these data, it was found that the mean value of Method A, $\bar{x}_A$ = 53.0, whilst that of Method B (referenced), $\bar{x}_B$ = 49.6. The mean paired difference $\bar{d}$ = 3.4 with its standard deviation s_d = 5.61. The sum of squares of differences SS_d=1043.

The t-value calculated from equation (1) was 3.033 which was larger than the critical value of t with (25-1) or 24 degrees of freedom was 2.064 for two-tailed test at 95% confidence level. This indicates that the final measurement of Method A was significantly different from that of the approved method B.

Table 1: Estimation of sampling bias by paired samples of a pesticide formulation consignment

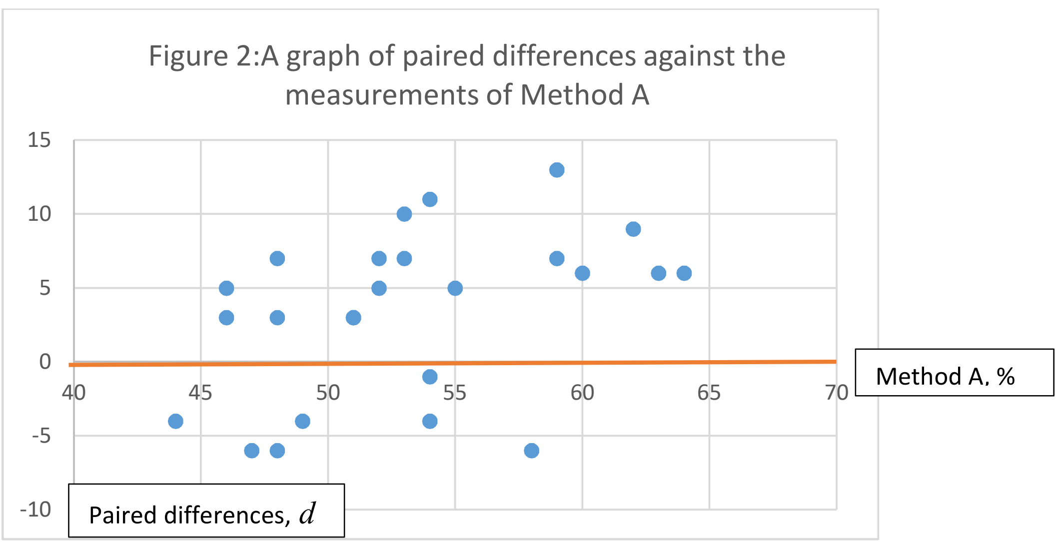
Target	Method 1	Method 2	Difference, d	d^2
1	54	55	-1	1
2	46	43	3	9
3	51	48	3	9
4	54	47	7	49
5	63	57	6	36
6	54	58	-4	16
7	44	48	-4	16
8	48	41	7	49
9	47	53	-6	36
10	52	47	5	25
11	53	43	10	100
12	64	58	6	36
13	48	45	3	9
14	55	50	5	25
15	59	46	13	169
16	62	53	9	81
17	49	53	-4	16
18	58	64	-6	36
19	52	45	7	49
20	60	54	6	36
21	54	43	11	121
22	51	48	3	9
23	48	54	-6	36
24	53	46	7	49
25	46	41	5	25

The outcome of the MS Excel® add-in Data Analysis tool on *t-Test: Paired Two Sample for Means* confirmed the above calculations obtained by the basic statistical principles:

t-Test: Paired Two Sample for Means

	Method A	*Method B*
Mean	53.0	49.6
Variance	30.667	36.417
Observations	25	25
Pearson Correlation	0.5336	
Hypothesized Mean Difference	0.0000	
df	24	
t Stat	3.0330	
P(T<=t) one-tail	0.0029	
t Critical one-tail	1.7109	
P(T<=t) two-tail	0.0057	
t Critical two-tail	2.0639	

If the differences between results of two sampling methods were plotted as a function of the corresponding concentration measurements of method A, the following graph (Figure 2) was generated:



Visually, it is also noted from the Figure 2 above that there is a significant bias as 18 out of 25 paired results were positive.

There also exists a trend which shows the absolute bias tends to increase with increasing concentration. See Figure 3.

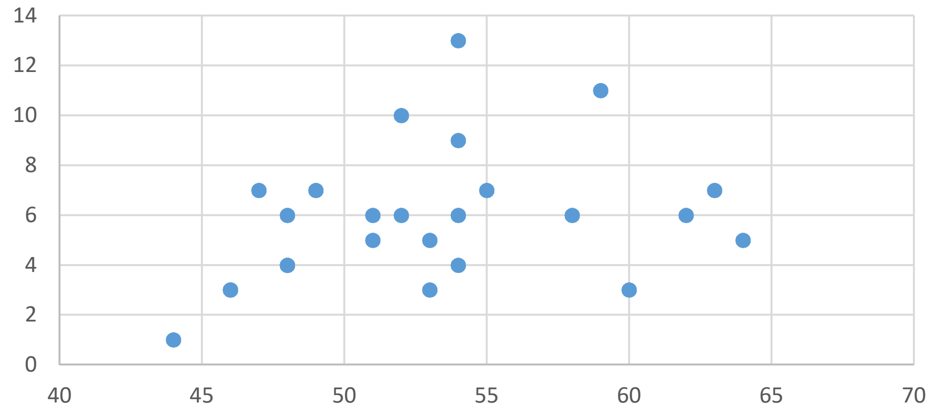


Figure 3: A graph showing the absolute differences against the measurements of method A

If the paired differences were to be randomly distributing around the zero dependent axis, (i.e. d = 0) in Figure 2, we can safely conclude that there is no significant bias and no suggestion of a dependence of bias on concentration.