A step-by-step ANOVA example on Sampling and Analysis

In conducting a hypothetical sampling exercise on one lot of palm kernel cake assignment for the analysis of oil/fat content (% w/w), let's say we have implemented a duplicate (double split design) sampling method, as recommended by Eurachem[1] and Nordtest[2] with the following strategy as an illustration.

Basically we have to make estimations of the variance of testing and variance of sampling:

1. From a population, we randomly select 6 target points for sampling, i.e. Target i with i = A, B, C, D, E and F

2. Split each Target sample into two equal portions, or randomly draw two samples from each Target, and designated them as Sample j, with labels S1 and S2

3. Carry out duplicate measurements (k =1 and 2) on each Sample j, giving two duplicate sets of test results, i.e. labelled as S1A1; S1A2; S2A1; S2A2

4. Conduct an analysis of variance (ANOVA) on the 6 Targets x 2 Samples x 2 duplicate analysis, or 24 test data, i.e. $x(i,1,1)$; $x(i,1,2)$; $x(i,2,1)$; $x(i,2,2)$

The test results of fat contents (%w/w) determined by the Soxhlet solvent extraction method are summarized in Table 1 below:

Table 1: The raw data of 6 Target points with duplicate samples and duplicate analyses

Target, i	S1A1	S1A2	S2A1	S2A2
%w/w	x_{i11}	x_{i12}	x_{i21}	x_{i22}
A	8.81	8.57	7.80	8.04
B	7.63	8.10	8.82	8.50
C	7.24	7.58	8.10	7.84
D	8.66	8.41	7.84	8.09
E	7.40	7.92	8.74	8.22
F	8.85	8.51	8.44	8.04

A. Estimating the variance of testing (within-sample)

1. Calculate the means of Targets A to F and the overall mean, using the general formula:

$$\bar{x} = \frac{\sum x_i}{n}$$

The mean values of these Targets are summarized in Table 2:

Table 2: Summary of calculated mean values of Targets A to F in S1, S2

Target, i	S1	S2
	$\bar{x}_{i,1}$	$\bar{x}_{i,2}$
A	8.69	7.92
B	7.87	8.66
C	7.41	7.97
D	8.54	7.97
E	7.66	8.48
F	8.68	8.24

The overall mean is: $\bar{X}_{total} = 8.17\% w/w$

(1) The variance of testing (within-sample) is estimated based on the *absolute* difference (also known as residual in statistics) of the data from each respective sample mean value (re: S1 and S2). For example, for Target AS1, we have:

$$x_{1,1,1} = 8.81,\ x_{1,1,2} = 8.57 \text{ and } \bar{x}_{1,1} = 8.69$$

$$D_{1,1(\bar{x})} = | x_{1,1,1} - \bar{x}_{1,1} | = |8.81\text{-}8.69| = 0.12$$

$$D_{1,1(\bar{x})} = | x_{1,1,2} - \bar{x}_{1,1} | = |8.57\text{-}8.69| = 0.12$$

It may have been noted that these two deviations are exactly the same, which should be the case.

(2) Therefore, the sum of squared differences of each target sample is calculated as:

$$D^2_{1,1(\bar{x})} + D^2_{1,1(\bar{x})} = 2 \times D^2_{1,1(\bar{x})} = 2 \text{ x } (0.12)^2 = 0.0288$$

The D's for the rest of the samples are calculated in the similar manner and tabulated the results in Table 3 below:

Table 3: Summary of sum of squares of Differences of each set of Target samples

Target, i	S1	S2
	2 x $D^2{}_{i1(\bar{x})}$	2 x $D^2{}_{i2(\bar{x})}$
A	0.0288	0.0288
B	0.1105	0.0512
C	0.0578	0.0338
D	0.0313	0.0313
E	0.1352	0.1352
F	0.0578	0.0800

(3) The total sum of squared differences in S1 and S2 (within samples), which is also known as sum of squared error SSE in testing, is calculated by summation of the sum of the squares of all the samples' D's:

$$SSE_{analysis} = 2 \times \sum_{i=1}^{6} \left[D^2{}_{i1(\bar{x})} + D^2{}_{i2(\bar{x})} \right] = 0.7816$$

(4) The degrees of freedom, $df_{analysis}$ is calculated from:

$$df_{analysis} = i\left[(j \times k) - j \right] = 6\left[(2 \text{ x } 2) - 2 \right] = 12$$

where, as indicated earlier,

(i) is the number of targets analyzed

(j) is the number of samples sampled in each target, and,

(k) is the number of repeated analysis performed in each sample

(5) The mean of squared differences in laboratory analysis, $MS_{analysis}$, is given by:

$$MS_{testing} = \frac{SSE_{analysis}}{df_{analysis}} = 0.7816/12 = 0.0651$$

(6) The variance of testing, $V_{testing}$, which is equal to $MS_{testing}$, is then calculated as :

$$V_{testing} = MS_{testing} = \frac{SSE_{analysis}}{df_{analysis}} = 0.0651$$

(7) The standard deviation and relative standard deviation (RSD) of testing are calculated from:

$$s_{testing} = \sqrt{V_{testing}} = 0.26\ \%\text{w/w}$$

$$RSD_{testing} = \frac{s_{testing}}{\overline{X}_{total}} \times 100\% = 3.12\%$$

B. Estimating the variance of sampling (between-sample)

To estimate the variance of sampling, the variance of overall measurement (sampling cum testing) in this sampling exercise has to be first evaluated.

(1) In each sampling Target point, we have taken two samples and run duplicate analysis on each sample for a total of 4 analyses.

So, the mean value of each Target, i, (two samples, 4 analyses) is calculated as:

$$\overline{X}_i = \frac{\overline{x}_{i1} + \overline{x}_{i2}}{2}$$

Each target's various mean values calculated are summarized in the Table 4 below.

Table 4: Summary of mean values of each Target for S1 and S2, and their respective overall mean

Target, i	S1	S2	
	$\overline{x}_{i,1}$	$\overline{x}_{i,2}$	$\overline{X}_i$
A	8.69	7.92	8.31
B	7.87	8.66	8.26
C	7.41	7.97	7.69
D	8.54	7.97	8.25
E	7.66	8.48	8.07
F	8.68	8.24	8.46

(2) Now, the difference between each of the sample mean and its respective overall mean $\overline{X}_i$ denoted as $D_{i(\overline{x})}$ is to be calculated by the following equation:

$$D^2{}_{i(\overline{x})} = \left(\overline{X} - \overline{x}_{i,1}\right)^2 = \left(\overline{X} - \overline{x}_{i,2}\right)^2$$

Hence, there are 4 such $D^2{}_{i(\overline{x})}$ for our consideration, namely:

$$\left(\overline{X} - \overline{x}_{1,1}\right)^2; \left(\overline{X} - \overline{x}_{1,2}\right)^2; \left(\overline{X} - \overline{x}_{2,1}\right)^2; \left(\overline{X} - \overline{x}_{2,2}\right)^2$$

(3) Therefore, the sum of squares of measurement, $SS_{measurement}$ (between-sample; sampling+analysis) is calculated according to:

$$SS_{measurement} = 4 \times \sum_{i=1}^{6} D^2{}_{i(\overline{x})} \quad = 2.7294$$

(4) The degree of freedom is calculated from the number of Targets, i, and the number of samples analyzed of each Target, j:

$$df_{sampling} = i \times j - i \quad = 6 \text{ x } 2 - 2 = 6$$

(5) The mean of squared measurement, $MS_{measurement}$ is:

$$MS_{measurement} = \frac{SS_{measurement}}{df_{measurement}} = \frac{2.7294}{6} = 0.455$$

(6) The variance of sampling, $V_{sampling}$, is given by:

$$V_{sampling} = \frac{MS_{measurement} - MS_{testing}}{2} = \frac{0.4549 - 0.0651}{2} = 0.195$$

Note: the denominator 2 refers to the number of samples involved in each Target point.

(7) It follows that:

the standard deviation of sampling = $\sqrt{V_{sampling}} = 0.44\% w/w$

and, $RSD_{sampling} = \frac{0.441}{8.17} \times 100 = 5.40\%$

C. Estimating the combined relative standard uncertainty of measurement

$$RSD_{measurement} = \sqrt{RSD^2{}_{sampling} + RSD^2{}_{testing}} = \sqrt{5.40^2 + 3.12^2} = 6.24\%$$

D. Reporting expanded uncertainty, U

Based on the $RSD_{measurement}$, we can then obtain a list of combined standard uncertainty, u and expanded uncertainty, U with a coverage factor of 2 against the values of % fat in the range covered as shown below:

Table 5: Corresponding u and U of fat contents (%w/w)

Given value %w/w	Standard u	Expanded U
7.50	0.47	0.94
8.00	0.50	1.00
8.50	0.53	1.06
9.00	0.56	1.12

E. Recommendation

It is recommended to have the above calculations carried out on an MS Excel® spreadsheet which can be made as a Template, so that the estimation of MU can be easily reviewed and updated as and when necessary.

F. References:

[1] Eurachem/CITAC Guide (2007) “Measurement Uncertainty arising from Sampling”

[2] Nordtest Technical Report TR 604 (2007) “Uncertainty from Sampling”