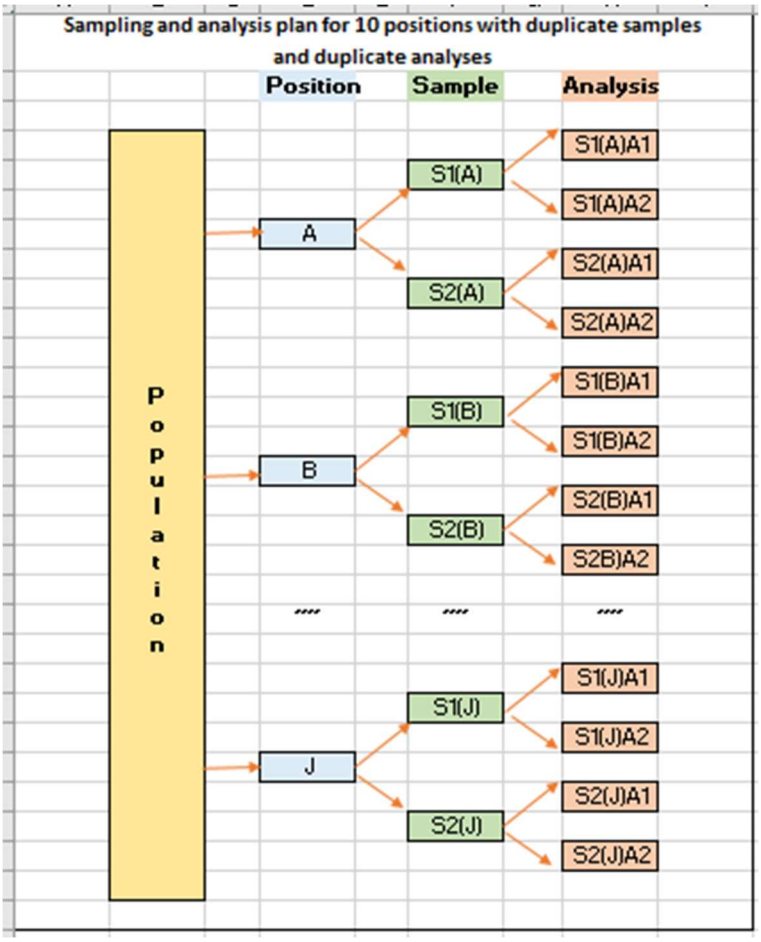
A worked example of measurement uncertainty estimation on non-homogeneous population

For sampling a non-homogeneous target population such as grain cargo, grainy materials or soil, random positions may be selected and split duplicate samples are taken with duplicate laboratory analysis carried out on each sample received. This approach will be able to address both sampling and analytical uncertainties at the same time.

Figure 1 below shows a sampling plan for 10 randomly selected positions with duplicate samples and duplicate analysis. So, a total of 10 positions x 2 samples x 2 analyses or 40 analytical data are now available for evaluation of its measurement uncertainty covering both sampling and analytical uncertainties.



For example, in an environmental soil investigation study of a chromium salt polluted land, the following duplicate analytical results of total chromium concentrations as Cr in mg kg^{-1}, analyzed by the acid-digestion / ICP-OES technique were obtained after taking duplicate samples from 10 randomly selected positions of the field as shown in Table 1:

Table 1: Total chromium Cr measurements

Position, i	S1-A1	S1-A2	S2-A1	S2-A2
	$x(i,1,1)$	$x(i,1,2)$	$x(i,2,1)$	$x(i,2,2)$
A	134	148	165	155
B	245	231	265	276
C	65	78	45	59
D	202	218	186	165
E	345	340	345	356
F	311	289	267	288
G	222	245	243	256
H	145	120	115	121
I	286	272	226	242
J	326	336	321	297

To estimate the measurement uncertainty, we need first of all carry out a two-factor (or two-way) analysis of variance (ANOVA), in order to study the data variations of within-sample, between-samples and between-positions.

Let's see how this can be done based on the basic ANOVA principles.

Within-sample variation

First, we estimate the analytical error.

In here, we have: number of positions, i = 10 with number of samples, j = 2 and number of repeated analysis, k = 2, and an overall mean of all the 40 data = 223.8 mg kg^{-1}.

The mean Cr results of each and every sample drawn are summarized in Table 2 below:

Table 2: Means of repeated analyses and sums of squares of duplicate samples

Target, i	S1	S2	S1	S2
	Mean i,1	Mean i,2	SSD(i,1)	SSD(i,2)
A	141	160	98	50
B	238	270.5	98	60.5
C	71.5	52	84.5	98
D	210	175.5	128	220.5
E	342.5	350.5	12.5	60.5
F	300	277.5	242	220.5
G	233.5	249.5	264.5	84.5
H	132.5	118	312.5	18
I	279	234	98	128
J	331	309	50	288

The Table 2 also shows the results of sum of squares of deviation (SSD) for all sample analyses. For example, the calculation of the sum of squares of deviation for Target A S1, expressed as SSD (i = A,1), is actually (134 – 141)^2 + (148 - 141)^2 = 49 + 49 = 98.

We may also use the Excel's spreadsheet function =DEVSQ(134,148) to get exactly the same answer.

Therefore, the total sum of squares of analysis error, SSE(analysis) = $\Sigma SSD(i,1) + \Sigma SSD(i,2)$ = 2616.5, with degrees of freedom in analysis, df (analysis), = (10 x 2 x 2) – (2 x 2) = 20.

It follows that the variance of analysis expressed as the *mean square* MS(analysis) = SSE(analysis) / df (analysis) = 2616.5 / 20 = 130.825.

The standard uncertainty of analysis, $u(\text{analysis}) = \sqrt{MS(analysis)}$ = 11.44, and the relative standard deviation (uncertainty) $\%RSD$(analysis) = (11.44 x 100)/ 223.8 = 5.11.

Between-samples variation

To study the between-samples variation, we must examine the variations amongst the mean values of the samples. Table 3 summarizes the mean values of these samples and the means of each location:

Table 3: Means of the duplicate samples and their sums of squares of deviation

Location, i	S1	S2	Mean of 1,2	$SSD(i)$
	$Mean\ i,1$	$Mean\ i,2$		
A	141	160	150.5	180.5
B	238	270.5	254.25	528.1
C	71.5	52	61.75	190.1
D	210	175.5	192.75	595.1
E	342.5	350.5	346.5	32.0
F	300	277.5	288.75	253.1
G	233.5	249.5	241.5	128.0
H	132.5	118	125.25	105.1
I	279	234	256.5	1012.5
J	331	309	320	242.0

Similarly, $SSD(i = \mathrm{A})$ can be calculated as $(141-150.5)^2+(160-150.5)^2 = 9.5^2 + 9.5^2 = 180.5$. We may also use the Excel's function =DEVSQ(141,160) to get the same answer.

Now, the sum of squares SS(between-samples) = j x ΣSSD_i = 2 x 3266.62 = 6533.25, where j = number of samples drawn at each location.

The degrees of freedom df(between-samples) = 10 locations x 2 samples – 10 locations = 10. Therefore, the mean square MS(between-samples) = SS(between-samples) / df(between-samples) = 6533.25 / 10 = 653.325.

It may be noted that the variance of between-samples as expressed by MS(between-samples) covers the sampling and analysis variances. Their relationship is as follows:

$$MS(between-samples) = k \times MS(sampling) + MS(analysis)$$

where k = number of repeats and in this case, k = 2.

Therefore,

$$Var(sampling) = \frac{MS(between-samples)-MS(analysis)}{k} = \frac{653.325-130.825}{2} = 261.25$$

and standard uncertainty of sampling, $u(sampling) = \sqrt{Var(sampling)}$ = 16.163.

The relative standard deviation (uncertainty) of sampling $\%RSD$(sampling) = 16.163 x 100/223.8 = 7.22

In summary, we have $\%RSD$(analysis) = 5.11 and $\%\,RSD$ (sampling) = 7.22

Estimation of measurement uncertainty

Hence, the combined standard $\%\,RSD$ (measurement uncertainty) = $\sqrt{\%RSD(analysis)^2 + \%RSD(sampling)^2}$ = $\sqrt{5.11^2 + 7.22^2} = 8.85$

The expanded measurement uncertainty of this exercise in terms of RSD is therefore 2 x 8.85 or 17.7%.

Table 4 shows examples of calculated measurement uncertainty of the determined Total Cr concentrations at different levels:

Table 4: Measurement uncertainty estimates based on %RSD found

Result mg kg^{-1}	Comb Std u	Expanded U
30	2.7	5.3
70	6.2	12.4
100	8.8	17.7
180	15.9	31.9
300	26.5	53.1
350	31.0	61.9

Important Note:

The above calculations covered only the random error aspects of the measurements. Sampling and analytical biases (i.e., systematic errors) if any, have not been considered here. If they were to be found important, then $u_{bias}(sampling)$ and u_{bias} $(anlaysis)$ estimated can be added as additional uncertainty components in the final approach.