Chapter 8

Determinate and indeterminate errors and cost

We saw in Chapter 5, Section 5.6 that there are a number of sources of error. The random or indeterminate errors are discussed in the requisite chapters 5, 7, 8 and later in 9. However, when it comes to determinate errors (see chapters 7 and 9), i.e. those associated with performing a process or operation that arise due to a fault in the way it is carried out, then there are three that can immediately be considered.

i) that due to the method employed, ii) that due to the instrumental technique used and iii) that due to the person involved. We will see in Chapter 9 that quality assurance, validation and peer review approaches can help us to identify these "systematic" errors that introduce "bias" and to some extent, address the determinate errors associated with the analytical approach or process.

Now, it is noted that the cost implication from failing to identify both the magnitude of the errors in the process and what is acceptable, can be financially disastrous. In terms of cost, it can often be the sample preparation and measurement steps that dominate the analytical process; hence, more samples that are measured will provide greater confidence in the values obtained but at a cost, both in time and money. The requirement of employing validated, quality assured procedures is not in question but an important consideration must always be the magnitude of the indeterminate errors. Section 8.2.1 introduces this very topic but one further approach within this area is that of considering the overall required level of confidence in the values obtained, the error associated with each step in the processes undertaken and the subsequent evaluation of the analytical approach to be taken, to meet these requirements.

For example, we saw in equation 8.1 that when considering the variance of the total process:

$$s^2_{tot} = s^2_{sam} + s^2_{sp} + s^2_{meas}$$

And that in many cases:

$$s^2_{sam} >> s^2_{sp} \geq s^2_{meas}$$

Therefore if the sampling variance is going to dominate the associated process errors then the acceptable error for the total analytical process (evaluated or identified from asking the right questions early on) will have a bearing upon the costs involved; i.e. the number and size of samples acquired, the number of samples to be prepared, the number of samples to

be measured using a particular analytical technique or methodology. Some analytical approaches are governed by regulations, legislation or guidelines and therefore will impose or dictate the requisite procedures. However, within the broader scheme of solving the analytical problem, the overall requirements of an evaluation within particular limits will still allow some flexibility of process to be cost effective; be it for the client and / or the laboratory!

Problem 1

A company that involves metal plating asked for its latest batch of solid waste material (two tonnes, destined for landfill) to be measured for its cadmium content. Many samples were taken and from these, five were selected at random to provide preliminary values of mean Cd content and standard deviation (SD). Calculations showed that the mean Cd value of 39.4 mg/kg had an error associated with sampling (n = 5) of 12 mg/kg, while that associated with the combined 'sample preparation and measurement' steps was 1.8 mg/kg; both expressed as their SDs.

Calculate by how much the <u>sampling error</u> must be improved in order for the overall (total) error to be less than 4 mg / kg

----

-

Number of samples taken and confidence

Table 8.1 in chapter 8 allows us to start the important cost debate. As was pointed out previously, each step involves a cost implication. As a result, it is instructive to evaluate as early on as possible just what is acceptable. For example, a farmer who requires a measure of the nitrogen and/or phosphorus levels in a 5 hectare field will expect a figure they can "deal with" (and not a debate on the mean values within a statistical set of variance limits at a particular confidence level!). If, as we have seen that the sampling step introduces the largest errors, then this effect can be reduced as the number of samples taken increases; a result of increasing the confidence associated with taking an increasingly 'representative' sample. The variance, however, is the square of the sample standard deviation, s^2_{sam} , and in order to reduce the RSD associated with a low number, n, of (particularly) heterogeneous samples taken, that number would have to increase as the square of the ratio of the RSDs.

If s_{rel} = the desired RSD and s_r = the RSD from a limited number of samples, then using equation 8.24 and rearranging we obtain:

$$n = [s_r / s_{rel}]^2\ t^2_{n-1}$$

As noted, t varies with the degrees of freedom (d.f) where d.f. = n-1, then for a given confidence level (95%), and for values of n-1 >15 ($t \leq 2.13$), we can <u>approximate</u> the equation to show;

$$n \cong 4.2\ [s_r / s_{rel}]^2$$

(as t^2 varies from factors ~4.5 down to ~4.0 for d.f. values of 15 up to 60). This approximation can be improved upon, if required, using the full calculation as shown in the example in section 8.3.2. for comparison.

Problem 2

You have guess-timated that 30 samples of soil will be required from a newly cleared 2 hectare field site at a local farm, in order to determine their levels of total 'nitrogen (N) and phosphorus (P)', in preparation for growing cereal. The values obtained will inform the farmer how much fertilizer will need to be added (which is an expensive process) and both you and the farmer have no previous values to go on! The 30 samples were finally acquired using a systematic, aligned grid technique, which took very little time (a few hours on site) and all were labelled and brought back within the morning. To turn the guess-timate into an estimate (using a statistical basis), you chose five out of the 30 samples (that had been quickly dried, ground and sieved) for further processing and measurement for N and P levels. By the end of the day, you know the following;

Sample No. ↓	Nitrogen (N); mg / kg	Phosphorus (P); mg / kg
3	5.3(3)	17.4(8)
7	7.1(6)	15.6(1)
14	4.4(7)	12.9(2)
21	3.9(2)	19.1(6)
28	9.2(1)	10.9(8)

Firstly: Calculate the mean, median, sample standard deviation and relative standard deviation (the latter both as a ratio factor and as a %) for both the nitrogen and phosphorus content of the soil samples.

OXFORD
UNIVERSITY PRESS

Secondly: Based upon these values, calculate at the 95% confidence level, how many samples need processing and measurement in order to provide, at least a relative standard deviation of 15% on the mean value for nitrogen and 10% on the mean value for phosphorus.

Thirdly: Calculate the size, as a radius in metres, of the circular (hot-spot) boundary associated with each sample (95% confidence), having taken 30 samples from the 2 Hectare field.

Problem 3

A 10 tonne batch of Kabuli chickpeas has been contaminated with another variety of chickpea, called Desi. This was a result of the Kabuli being accidentally added to a silo already holding a quantity (100 kg) of the Desi variety. Unfortunately, it was thoroughly mixed in the silo before the contamination issue was identified. The batch can still be sold as long as the contents are known to a suitable level of precision. Assuming the particle size and density of the two varieties of chickpea are the same (5 mm dia; 1.20 g / cm^3) what mass of sample should be taken in order to ensure a level of precision of 1% identification on the contaminant concentration?

Hint: See section 8.3.1

Heterogeneity and the sampling of multi-component solids

The examples shown in Chapter 8 begin to demonstrate the range of approaches that can be taken to identify errors in the sampling of solids and in tackling some of their associated complexities. Example 8.2 introduces one approach to sampling a heterogeneous material such as an ore. Of course, the focus may well be upon one component within the material (the selected mineral within the ore matrix) which is why **equation 8.9** is presented in the way it is. Once the individual, contributing values for this equation are identified for a particular material (usually from some preliminary studies of ores), these can be pasted into the equation to provide an estimate of the parameter under investigation.

Problem 4

A batch (25 tonnes) of a crushed and graded ore of sphalerite [Zn(Fe)S] in quartz, is to be sampled. This is to provide a measure of the Zn content with a precision of ±0.05% as part of its assay to a confidence level of 95.4% (≡ ẍ± 2SD). The crushed ore's nominal particle size is 5 mm (some 95% of the ore passes through this sieve size) while the grain size or liberation size (from microscopy) for the sphalerite in this batch is ~ 0.1 mm (100 µm). The ore, from initial testing has a Zn content of ~ 3.33% and is basically the mineral sphalerite in a quartz matrix. Given that sphalerite has a density range of 3.9 up to 4.1 g / cm^3 (dependent upon its Fe content which is <u>very low in this case</u>), you have been asked to assume the density as 4.1 g / cm^3 while that for quartz is 2.65 g / cm^3. The shape factor of the sphalerite is ~0.5 because its crystalline morphology from microscopy is tetragonal-like, while it displayed a broad particle size range, so its PSD factor is ~ 0.25.

The composition factor, 'c' is based upon using either,

i) the approximation formula of:

overall sample density (fraction of sphalerite x density$_{[Zn(Fe)S]}$ + fraction of quartz x density$_{Quartz}$) all divided by the concentration for the mineral [Zn(Fe)S] expressed as a fraction;

or

ii) upon the preferred formula below*:

$c = [(1-a) / a][(1-a)\rho_m + a\rho_g]$;

where a = fraction of required mineral in the ore

ρ_m = density of required mineral

ρ_g = density of the matrix material

The liberation exponent 'b' from studies of this ore type, has a value of ~ 0.5. Given this information, how much ore sample needs to be taken from this batch in order to provide the desired sampling accuracy stated above?

* See reference stated in 'Note 4' to equation 8.9 in chapter 8.

Heterogeneity and the Benedetti-Pichler equation

It is worth reflecting on what we have seen in the previous on-line example and in section 8.2 of chapter 8. A heterogeneous solid material can be represented, and hence modelled, by the following:

i)   The number of individual 'components'; P_a, P_b, P_c, P_d….

ii)   The distribution of each within the solid matrix; P_a within $P_b + P_c + P_d$ +….; P_b within $P_a + P_c + P_d$ +….; P_c within $P_a + P_b + P_d$ +….etc.

iii)   The density (ρ) of each component ρ_a, ρ_b, ρ_c,…

iv)   The particle size distribution (PSD) of each component, $(PSD)_a$, $(PSD)_b$, $(PSD)_c$ ….etc. within the matrix

If, for example we were to consider an analyte distributed between two components as the next step-up in complexity then using the binomial approach as a basis (equation 8.13), one constructed model[‡] can show the following relationship:

$$s^2 = [\rho_a \rho_b / \rho^2]^2 \, (P_a - P_b)^2 \, [p(1-p) / n]$$

Here, having a sample with only two components, A and B, whose densities are ρ_a and ρ_b and where the analyte of interest is distributed between these two components, their contents being P_a and P_b, then if the sample has p fraction of A type particles and (1-p) fraction of B type particles, the above equation provides the relationship between the number of particles sampled, n, and the absolute sampling variance s^2. The average density of the sample is given by ρ and in this case the components A and B have the same particle size.

The equation demonstrates some pertinent points. The section $(P_a - P_b)$ in the above reflects the level of heterogeneity of a sample. It is at its most heterogeneous where either P_a or P_b contains all of the analyte and the other component contains none. As $P_a \rightarrow P_b$ in terms of content then the level of heterogeneity and the variance decreases until the point where $P_a = P_b$ and the sample is homogeneous; resulting in the theoretical sampling variance being zero. The equation also shows (as we saw in chapter 8) that the variance decreases as i) the number of particles sampled increases and ii) the fraction $p \rightarrow 0$ or $p \rightarrow 1$. Of course if p = 0 or 1 then there would only be one type of component present and the sample would therefore be homogeneous.

If, however we note that our relative standard deviation, s_r , as a fraction, is given by:

$$s_r = s / P$$

where **P** = overall concentration in the analyte in the sample mixture then substituting for '**s**' in the above, our equation becomes:

$$s_r{}^2 = [\rho_a\, \rho_b\, /\, \rho^2]^2\, (P_a - P_b\, /\, \textbf{P})^2\, [p(1\text{-}p)\, /\, n]$$

which is a variation of the better known 'Benedetti-Pichler' equation[‡].

[‡] There are numerous up-to-date papers that utilise this equation but the original is from: Benedetti-Pichler, A. A. In *Physical Methods in Chemical Analysis,*
Vol. 3; Berl, W. G., Ed.; Academic: New York, 1956; pp 183–217.

Problem 5

A suitable quantity of a mixed sphalerite [Zn(Fe)S] ore (total, 25 tonnes) is required for analysis. The graded ore comprises two loads or "lots" (15 tonnes and 10 tonnes) of sphalerite in a quartz matrix, that have each been partially ground, sieved (to the same size) and then carefully mixed. Initial investigations indicate that the 15 tonne 'lot' contains ~20 % as Zn while the 10 tonne 'lot' contains ~10 % as Zn. Given that the molar masses of Zn and Zn(Fe)S are 65.38 g and 97.44 g respectively and the densities of Zn(Fe)S and Quartz (SiO_2) are 4.10 and 2.65 g / cm^3 respectively, calculate:

1) The number of particles required in your sample, in order to allow a precision of ± 1 % relative standard deviation (expressed here as a %).

2) The mass of sample required in order to meet the above precision based upon a mean particle size diameter of 10 mm from analysis. You may assume the particles to be spherical.

3) What number of particles and mass of sample would you need to take, to increase the confidence level in your sample to 95.4 % (2xSD)?

Problem 6

A calcite ($CaCO_3$) ore is found from preliminary analysis to contain dolomite ($CaMgCO_3$) as a contaminant, down at ~ 0.86 % (8600 mg / kg) concentration.

1) How much sample (by mass) would be required in order to measure the dolomite concentration to a level of precision of 1.0 % RSD, given that the sample density is 2.71 g / cm^3 and the ore particles have a mean diameter of 5 mm?

2) The mass of sample calculated from part 1) above, required to achieve the precision stated, is more than can be readily handled. What mean particle size should the ore be ground down to, in order to take back a more handle-able 5 kg sample to the laboratory which will achieve the same level of precision?

Problem 7 "How many samples of Tablets to take"

If we return to the Sampling Problem 4, in our on-line Chapter 2 additions, we were asked to consider formulating a sampling plan, for obtaining samples of tablets (containing our OTC drug) in blister packets from our production line.

Given that you are required to know at the 95% confidence level the drug content (value in mg) of the tablets to a relative standard deviation of ± 0.5% and that the previous production line runs presented values of 510 mg as the mean active drug content with a standard deviation of 10.2 mg, how many samples would you need to take?