

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:

$$s_{tot}^2 = s_{sam}^2 + s_{sp}^2 + s_{meas}^2$$

And that in many cases:

$$s_{sam}^2 \gg s_{sp}^2 \geq s_{meas}^2$$

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!

Feedback on problem 1

Given from equation 8.1 that the variances are additive, then

$$s_{tot}^2 = s_{sam}^2 + s_{sp}^2 + s_{meas}^2$$

By inserting the requisite values; s_{tot}^2 is:

$$(12)^2 + (1.8)^2 = 14(7) \text{ (mg / kg)}^2$$

$$\text{and } s_{tot} = 12.(1) \text{ mg / kg}$$

If a total error of less than 4 mg / kg is required then the equation shows:

$$(4)^2 = s_{sam}^2 + (1.8)^2 ;$$

Given that the $s_{sp}^2 + s_{meas}^2$ steps will vary very little and $s_{sam}^2 \gg s_{sp}^2 \geq s_{meas}^2$ in comparison. Then

$$s_{sam} = 3.5(7) \text{ mg / kg}$$

Therefore the improvement in s_{tot}^2 down to < 4 mg / kg requires s_{sam} going from 12 mg / kg down to 3.5(7) mg / kg. This improvement would only be acquired from the processing and measurement of many more samples; as demonstrated in the later Problem 2 to be tackled.

Feedback on Problem 2

Using our statistical formulae from Chapters 7 and 8, and the equation discussed;
i.e. $n \cong 4.2 [s_r / s_{rel}]^2$

We obtain the following from our data.

Statistical Evaluation	Nitrogen	Phosphorus
Mean [mg / kg]	6.0(2)	15.2(3)
Median [mg / kg]	5.3(3)	15.6(1)
Std Dev [mg / kg]	2.1(7)	3.3(2)
RSD (factor)	0.36(0)	0.21(8)
% [RSD factor x 100]	36.(0)	21.(8)
Required RSD %	15	10
$[s_r / s_{rel}]$	2.4(0)	2.1(8)
$[s_r / s_{rel}]^2$	5.7(6)	4.7(5)
$\sim n$	24.(2)	20.(0)
minimum value of $\sim n$	24 to 25	19 to 20
Full calculation (and minimum) value for n	23	19

It is noted that the nitrogen and phosphorus values from the 5 initial samples processed have broad range RSDs of 36% and ~22%. The factor $[s_r / s_{rel}]$ in our equation for calculating 'n' means that we can use our relative standard deviations either as a factor or as a %. This ratio demonstrates the improvement required for our nitrogen and phosphorus measurements (2.4 x and 2.2 x respectively), while the minimum whole number of samples to be processed for each, in addition to the 5 already processed, demonstrates 20 more are required for nitrogen and 15 more for phosphorus. The values for n overall are also in the range 20 to 30 and therefore our approximation is acceptable for 't'. Using equation 8.24 and Figure 8.1 provides us with the full calculation 'n' values, as shown in the table of results for comparison.

Of course, with at least 25 samples to prepare, it may be possible to measure both the N and the P on each processed sample or it may be that two separate sample preparation techniques would be required, one for 20 samples and one for 25 samples. As previously noted, in terms of cost, it can often be the sample preparation and measurement steps that dominate the analytical process; hence, the more samples measured will provide greater confidence in the values obtained but at a cost, both in time and money.

To calculate the size of the circular boundary around each sample, we use equation 2.2 and then 2.1, where $n = 30$ and area $A = 20,000 \text{ m}^2$. The grid size is therefore 25.(8) m and the radius of the circular boundary is 15.(2) m.

What is “acceptable” and “fit for purpose”

Finally, on this topic, it is noted that a farmers requirements and “what is acceptable” will not necessarily be the same as that of, say a parliamentary committee on food and nutrition. While the committee would require a summary evaluation that they can (intellectually) deal with, they will also expect for public safety, a level of rigorous evaluation according to statutory requirements that have included a full statistical approach with error limitations to provide an accepted level of confidence in any values they set.

Overall, It is important that the time and effort expended in achieving ‘the required’ or ‘an acceptable’, level of confidence values and hence the costs associated in this endeavour are all considered. An old expression for such a process is: “chemist, be aware and beware of what you are asked for!”.

Feedback on Problem 3

As the two different varieties of chickpea have the same particle size and density, then this is a particle number problem. Hence, we can use equation 8.16. to calculate the first part of the problem.

Here, $p = 100 / 10100$; because 10 tonne = 10,000 kg and the contaminant is 100 kg.

Therefore $p = 0.0099(0)$ and our equation becomes:

$$n = [1 - 0.0099(0)] / 0.0099(0) \times (0.01)^2$$

$n = 1,000,000$. Chickpea particles should be sampled

It is noted that the two varieties of chickpea are actually different colours and therefore within the 1,000,000 particles sampled, there will be ~ 9,900 identifiable Desi chickpeas. Rather than subject an unfortunate assistant to counting out the 10^6 particles, thankfully an automated identification and counting instrument could be used or better still, the second part of the calculation (below) can be employed.

Given that each particle is 5 mm dia. and 1.20 g / cm^3 density, then using equation 8.17, the mass to take, W , would be:

$$W = n\rho V$$

Where ρ = density and V = volume ($\pi d^3/6$)

Substituting, we have

$$W = [1,000,000. \times 1.2 \times 3.142 \times (0.5)^3 / 6]$$

$W = 78,600 \text{ g}$ or $78.(6) \text{ kg}$. should be sampled to provide the level of precision on the concentration of the contaminant.

Feedback on Problem 4

In Chapter 8, equation 8.9 is our starting point. Here we see that for most cases:

$m_L \gg m_S$ and we know that our m_L , (25 tonnes) when expressed in grams is 2.5×10^7 ; hence $1 / m_L$ in the equation is very, very small (4×10^{-8})

So our equation to use is 8.10

The description above provides all the information we require but one factor that we need to calculate first is the variance, s^2 via the standard deviation (SD) 's'..

It states in the problem: "...a precision of $\pm 0.05\%$ as part of its assay to a confidence level (CL) of 95.4% ($\equiv \bar{x} \pm 2SD$). The 0.05% is the absolute SD precision on the ~3.3(3) % Zn concentration; i.e.; $\bar{x} \pm 0.05 \%$. Gy's formula (equation 8.10) refers to

the relative variance, hence at the 95.4% CL ($\equiv \bar{x} \pm 2SD$) stated, the variance becomes:

$$2.0 \times s = 0.050\% / \%Zn \text{ content of ore}; = 0.05 / 3.3(3) = 0.015(0)$$

$$s_r = 0.015 / 2.0 = 0.0075(0); s_r^2 = 0.000056(3)$$

The values for K and for d_n^3 will need to be calculated and here we keep all units as grams or cm or cm^3 .

From above, $d_n = 0.5 \text{ cm}$, so for $d_n^3 = 0.125 \text{ cm}^3$.

$K = f \times g \times c \times l$ and so from above: $f = 0.5$; $g = 0.25$;

Now c can be calculated by two approaches;

where the approximation method i) uses:

Sample density / Concentration of mineral as fraction

The overall 'sample density' was shown above to be:

fraction of sphalerite \times density $_{[\text{Zn}(\text{Fe})\text{S}]}$ + fraction of quartz \times density $_{[\text{Quartz}]}$

Now the Zn content is 3.3(3)% and sphalerite is expressed as $\text{Zn}(\text{Fe})\text{S}$, so if the Fe content is very, very low then the mineral approaches ZnS :

Zn = RMM of 65.38 and S = RMM of 32.06. By proportion, if the Zn content is 3.3(3)% then this is equivalent to $\sim 5.0\%$ (4.9(7)%) $\text{Zn}(\text{Fe})\text{S}$ as a maximum. The fraction of sphalerite is therefore 5/100 and the fraction of quartz is 95/100. Given that the densities are 4.1 and 2.65 g / cm^3 for sphalerite and quartz respectively, the overall sample density is:

$$[0.05 \times 4.1] + [0.95 \times 2.65] = 2.7(2) \text{ g} / \text{cm}^3.$$

The concentration of the mineral sphalerite expressed as a fraction is 0.05, so that the approximation to factor 'c' becomes:

$$2.7(2) / 0.05 = \sim 54.(4)$$

The other method, ii) is where $c = [(1-a) / a][(1-a)\rho_m + a\rho_g]$;

Here, a = fraction of required mineral in the ore (from above = 0.050)

As calculated, if Zn = 3.3(3)%, then Zn(Fe)S = 4.9(7)% = $\sim 5.0\%$; from 5/100 as fraction).

ρ_m = density of required mineral Zn(Fe)S = 4.1 g / cm³

ρ_g = density of the matrix material (Quartz, SiO₂ = 2.65 g / cm³).

Substituting the values into the equation, then $c = 76.(5)$; which is considered a more accurate measure and will be used.

The final factor is the Liberation factor ' f '. This is expressed as:

The grain size, d_l for liberating the Zn(Fe)S mineral sphalerite (= 0.010 cm) divided by the nominal ore size, d_n (= 0.50 cm), all raised to the power of 'b', the liberation exponent, stated to be 0.5 in the question.

Hence, the Liberation factor ' f ' = $[0.01 / 0.5]^{0.5} = 0.14(1)$;

Solving for m_s in equation 8.10 we have:

$$m_s = 0.5 \times 0.25 \times 76.(5) \times 0.141 \times (0.125) / s_r^2,$$

If $s_r^2 = 0.000056(3)$ then the mass, m_s is:

2994 g or 3.0 kg

Considering the significant figures, then a 3.0 kg sample should allow sufficient particles of the ore to provide an assay of the Zn content to a precision of $\pm 0.05\%$ (0.015 relative for 1 SD; 0.0075 relative for 2SD etc.) at a confidence level of 95.4%

Feedback on Problem 5

Rearranging the relevant equation, we have 'n', the number of particles:

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

From the information given, we know that 60 % of the mixed ore (15 tonnes; from 15 / 25) contains Zn at a concentration of 20 %; while 40 % of the total mixed ore (10 tonnes; from 10 / 25) contains Zn at a concentration of 10 %. Therefore we can identify the following:

$$P_a = 20 \%$$

$$P_b = 10 \%$$

$$p = \text{fraction of sphalerite in ore A particles; } = 60 \% / 100 = 0.60$$

$$1 - p = \text{fraction of sphalerite in ore B particles; } = 40 \% / 100 = 0.40$$

We can calculate 'P', the total quantity of Zn based on these:

$$P = [0.6 \times 20 \% \text{ Zn} + 0.4 \times 10 \%] = 12 + 4 = 16 \% \text{ Zn in the total mixed ore}$$

Next, we can calculate the densities of the different component particles A and B.

First we need to express the % Zn as % sphalerite [Zn (Fe)S] in the two components as the density is based upon the compounds present..

Using the molar masses, the % Zn $\times 97.4(4) / 65.3(8)$ in each case gives us:

For A

$$20 \times 1.49 = 29.8(1) \% \text{ sphalerite in component A}$$

And

$$100 - 29.8(1) = 70.1(9) \% \text{ quartz in component A}$$

If ρ_a = density of A particles,

$$\begin{aligned} \text{this makes our density for A} &= 0.298(1) \times 4.10 + 0.701(9) \times 2.65 \text{ g / cm}^3 \\ &= 3.08 \text{ g / cm}^3 \end{aligned}$$

And for B

$$10 \times 1.49 = 14.9(0) \% \text{ sphalerite in component B}$$

And

$$100 - 14.9(0) = 85.1(0) \% \text{ quartz in component B}$$

If ρ_b = density of B particles,

$$\text{this makes our density for B} = 0.149 \times 4.10 + 0.851 \times 2.65 \text{ g / cm}^3$$

$$= 2.86(6) \text{ g / cm}^3$$

If ρ = overall density of mixed ore then this will be given by:

$$60 / 100 \times 3.08 + 40 / 100 \times 2.86;$$

$$\rho = 2.9(9) \text{ g / cm}^3$$

The value of S_r is obtained from the information that we require a precision of $\pm 1\%$ relative standard deviation; S_r when given as a fraction is $1.0 / 100 = 0.010$

To calculate n , the number of particles, we now insert the relevant numbers into the rearranged equation;

$$n = [3.08 \times 2.87 / (2.99)^2]^2 (20 - 10 / 16)^2 [0.6(1-0.6) / (0.010)^2]$$

$$n = [0.977] [0.390] [2400]$$

$$n = 917$$

Therefore in answer to part 1) at least 917 particles of $\sim 1 \text{ cm}$ mean particle diameter will need to be sample to ensure the level of precision required.

This number may not seem very large and this is a result of the values for the various components associated with A and B having only a relatively small difference. The material is therefore not particularly heterogeneous and as we noted above, as the various values for the two components, A and B, approach one another, the less material overall has to be sampled for a given level of precision.

ρ_a , ρ_b , and ρ

P_a , P_b and P

p and $(1-p)$

It was noted in the question that the sample was also partly ground and sieved and well mixed (to reduce its heterogeneity).

For part 2) the relationship between number and mass is dependent upon the volume and the density. For a spherical particle this is given by equation 8.17 in chapter 8.

Mass = $n \rho V$; where ρ from the above calculation is 2.99 g / cm^3 and $V = \pi d^3 / 6$ with 'd' is given as 10 mm (= 1 cm)

Therefore from $917 \times 1 \text{ cm}$ particles of density 2.99 g / cm^3 , we have:

1440 g; if rounded up to cover various approximations then a 1.5 to 2 kg sample should be sufficient.

For part 3) we note that the equation

$$s_r = s / P$$

provides the relationship between the absolute standard deviation, s , and the relative standard deviation, s_r , both given as a fraction for a mean value of 'P'. The value of s_r can also be considered as given with regard to a level of confidence ($\pm 1 \text{ SD}$, $\pm 2 \text{ SD}$, $\pm 3 \text{ SD}$ etc.) on the value of 'P' ($P \pm zs$, where 'z' refers to the number of standard deviations from the mean value of 'P'). Hence, the relationship can be shown as:

$$s_r = zs / P$$

In order to increase the confidence level on the value of 'n' and on the mass of material taken, to 95.4 % or $\pm 2 \text{ SD}$, the value of s / P now = $s_r / 2$. The RSD of 1 % (factor of 0.01) asked for in the question now becomes 0.5 % or expressed as a factor, 0.005. In the full equation this changes the value of 'n' by 2^2 (as s_r^2 is used in the full equation). Hence, the number of particles becomes $4 \times 917 = 3,670$ and the mass of sample becomes 5,760 g or $\sim 5.8 \text{ kg}$.

Feedback on Problem 6

- 1) This is a particle number problem that can be converted into a mass using equations 8.16 and 8.19 from chapter 8.

Here, $p = 0.86 \% / 100 = 0.0086$ and $s_r = 1 \% / 100$; therefore:

$$n = (1 - 0.0086) / [0.0086 \times (0.01)^2]$$

$n = 1,152,800$ particles of the calcite ore

Now, the density of the ore ' ρ ' = 2.71 g / cm³ and the mean particle diameter is 5 mm (0.5 cm); therefore:

$$W = 1152800 \times 2.71 \times 3.142 \times (0.5)^3 / 6 \text{ g}$$

and

$W = 204,500$ g; hence some 205 kg would cover the quantity required for a precision of 1% RSD on the mean dolomite value.

- 2) If you only wish to bring back 5 kg to the laboratory then equation 8.19 can be re-arranged to show:

$$d^3 = 6W (\rho s_r^2) / \pi \rho (1-p)$$

Given that all other values are known, then substituting we have:

$$d^3 = 6 \times 5000 [0.0086 \times (0.01)^2] / 3.142 \times 2.71 \times (0.9914)$$

$$d^3 = 0.00306$$

$$d = 0.145 \text{ cm (1.45 mm)}$$

This shows us that our 5 kg ore sample will contain the same number of particles as before ($n = 1,152,800$ particles of the calcite ore) when the mean particle diameter is reduced from 5 mm down to 1.45 mm,

Feedback on Problem 7

If we look to Section 8.3.2, the equations presented allow us to calculate the number of samples to take in order to allow us the confidence to state that a value is within certain limits. Considering equation 8.24, we can state that in this case, $s_{rel} = \bar{X} - \mu ; = 0.005$ (from 0.5%) and $s_r = 10.2 / 510 = 0.02$ (from 2.0% RSD shown by production line). Then:

$$(0.02)^2 / (0.005)^2 = 16.$$

And using the equations from 8.24 we see that this value of $16 = (d.f. + 1) / t^2$

We know from Chapter 7, that there is a relationship between number of samples n , the degrees of freedom (d.f, where $d.f. = n - 1$) and the t -test which provides us with a t factor value for a given confidence level (CL). We normally use the 95% CL. However we also know that as n changes, so does the t value. Therefore, we can use the graphical plot shown in Figure 8.1 of Chapter 8 to obtain a value for d.f directly* and hence n ;

From Figure 8.1 we can interpolate by placing the value of 16 on the y-axis, to provide us with the degrees of freedom (d.f.) on the x-axis . Noting that $d.f = n - 1$, where $n =$ number of samples.

The graph shows $d.f. = 62$

As $n-1 = d.f.$,

then $n = 63$;

Therefore we would need to take at least 63 samples, to provide us (at the 95% confidence level) with a relative standard deviation of 0.5 % on our mean value. In our on-line problem from Chapter 2 for this scenario, we estimated that 100 samples might be a good starting point. Given the production run's statistical values, this actually allows us to calculate and therefore improve our estimate.

* Other methods have an iterative approach, often requiring multiple steps before the value for n is obtained.