Chapter 2

It is important to note that in any sampling process, your sampling plan identifies the required health and safety measures to incorporate (as stated in chapters 1 and 2). A number of guidance documents are available and one which focuses on fieldwork is:

The BGS + NERC Safety fieldwork:
https://nerc.ukri.org/about/policy/safety/procedures/guidance-fieldwork/

Problem 1

To demonstrate the effect of particle size distribution on taking a representative sample:

Part (a)

For this demonstration, you will require the following quantities of three different sized, spherical plastic beads.

i) 10,000 x 2 mm dia.
ii) 600 x 5 mm dia.
iii) 80 x 10 mm dia.

You will also require a 10 mL plastic beaker and a clear, colourless plastic tube, around 70 to 80 mL in volume and approx. 100 to 120 mm long, 25 to 30 mm diameter which can be sealed off at the ends. Disposable plastic centrifuge tubes with screw top caps are available in this size range. Place a mark along its side with an indelible marker to identify the volume 60 mL.

It is noted that these quantities of beads shown above have a similar bulk density, i.e. they have a very similar mass, when each quantity occupies ~60 to 70 mL in their unpacked state. This bulk effect can be observed by pouring the 2 mm beads into the plastic tube and up to the 60 mL mark. This quantity of beads can then be weighed on a tared balance. This value is noted and the same procedure is repeated with the 5 mm and then the 10 mm beads. If the same material (e.g. same plastic) is used for all three sizes of beads then the mass of these 60 mL quantities will be similar.

Next,

a) Place 12 of the 5 mm beads in the tube and then pour the 2 mm beads into the tube up to the 60 mL mark. Place the cap on the tube to seal it and then rotate the tube, end over end and side to side slowly and carefully, in order to disperse the 5 mm beads and to attempt a homogeneous mix. Do this for 20 seconds.

b) Carefully pour the contents (without shaking) into the 10 mL plastic beaker and when 10 mL has been added, check the contents of this first “batch” for the number of 5 mL beads. Note this number and then place all these beads to one side.
c) Without shaking, carefully pour another 10 mL volume of beads from the tube into the 10 mL beaker and again check the contents of this second “batch” for the number of 5 mL beads. Note this number and then place these beads to one side.
d) Without shaking, carefully repeat step c) four more times. This will give you six batch values, each with its own number of 5 mm beads.

Compare your individual results for each batch from your experiment in a table, and then compare between each experiment undertaken by the class. What can you say at this stage about your values within your own experiment and between the experiments in the class overall?

Part (b)

Segregation effects. Different sized particles in a tube to show the effect of particle size segregation.

Using the apparatus described in Part (a), add five of the 10 mm beads to the 60 mL tube and then pour in the 2 mm beads up to the 60 mL mark and cap the top. The 10 mm beads are now at the bottom of the tube and the 2 mm beads above, up to the 60 mL mark.

Holding the tube upright at all times for this next step of the experiment, with its base firmly on the table, gently tap the outside of the plastic tube from the bottom to about half way up its side and back down to the bottom on a continuous cyclic basis, allowing the contents to be shaken and vibrated. Continue this tapping until you see any difference to the beads at the surface of the tube (this can take a while!). Describe what you see.

Continue to tap the upright tube until no further change occurs at the surface. Describe what you now see.

Keeping the tube up-right and steady, carefully un-seal the top of the tube and gently pour the top 10 mL of the tubes contents into the 10 mL beaker. Describe what you see in the beaker in terms of bead population and particle size.

Example Case Study - to demonstrate the effect of distribution on sampling:

One further example of the effects of distribution and the possible errors that can result from “sampling”, particularly as a result of the number of samples taken, is shown by the “cube” scenario. Here, a cube of material (a “lot” of material) weighing 1.0 kg is theoretically divided up into 32 equally sized “test portions” or sub-samples for the laboratory, where each test sample is itself, in the shape of a cube but all sub-samples remain in their original position within the larger cube. As a whole, the 1.0 kg lot contains 32 mg of copper. Each sub-sample

weighs 1 / 32 kg and would contain 1 mg of Cu, if the distribution of the analyte of interest (i.e. copper) is perfectly homogeneous. Any of the 32 cubes would reflect a representative sample, e.g. If four of the 32 sub-samples are taken, say from the corners of the “lot”, the total amount of Cu found would be 4 mg present in $1/8^{th}$ of the mass of the total “lot”. Our calculation is therefore:

4 mg Cu per $1/8^{th}$ kg, or 32 mg Cu / kg as a concentration

Now consider the effect produced when the distribution is based upon the following:

Half (16) of the sub-samples each contain 1.5 mg Cu and the other 16 sub-samples contain 0.5 mg of Cu. The total amount of Cu present in the 32 cubes is still 32 mg and therefore the concentration overall is 32 mg / kg as before. The distribution is relatively homogeneous such that the sub-samples are uniformly spread, throughout. However when sampling these cubes, the accuracy in terms of the overall concentration determined (shown as % recovery of the overall mean per cube = 1 mg) and the precision (the reproducibility from the sampling) will depend upon the number of samples taken (n = number of cubes). We could therefore produce the following possible sub-sampling effects as we increase the number of samples taken and options in recovery:

Number of samples taken (n)	Value in mg of Cu	Mean Value, mg total value/n	Recovery as a %	Precision
1	0.5	0.5	50	
1	1.5	1.5	150	
2	0.5+0.5	0.5	50	good
2	0.5+1.5	1.0	100	v. Poor
2	1.5+1.5	1.5	150	good
3	0.5+0.5+0.5	0.5	50	v. Good
3	0.5+0.5+1.5	0.833	83.3	poor
3	0.5+1.5+1.5	1.166	116.6	poor

3	1.5+1.5+1.5	1.5	150	good
4	0.5+0.5+0.5+0.5	0.5	50	v. good
4	0.5+0.5+0.5+1.5	0.75	75	fair
4	0.5+0.5+1.5+1.5	1.0	100	v. poor
4	0.5+1.5+1.5+1.5	1.25	125	fair
4	1.5+1.5+1.5+1.5	1.5	150	v. good
5	0.5+0.5+0.5+0.5+0.5	0.5	50	v. good
5	0.5+0.5+0.5+0.5+1.5	0.7	70	v.fair
5	0.5+0.5+0.5+1.5+1.5	0.9	90	poor
5	0.5+0.5+1.5+1.5+1.5	1.1	110	poor
5	0.5+1.5+1.5+1.5+1.5	1.3	130	v.fair
5	1.5+1.5+1.5+1.5+1.5	1.5	150	v. Good
6	0.5+0.5+0.5+0.5+0.5+0.5	0.5	50	v. good
6	0.5+0.5+0.5+0.5+0.5+1.5	0.666	66.6	fair
6	0.5+0.5+0.5+0.5+1.5+1.5	0.833	83.3	poor
6	0.5+0.5+0.5+1.5+1.5+1.5	1.0	100	v. poor
6	0.5+0.5+1.5+1.5+1.5+1.5	1.166	116.6	poor
6	0.5+1.5+1.5+1.5+1.5+1.5	1.333	133.3	fair
6	1.5+1.5+1.5+1.5+1.5+1.5	1.5	150	v. good

In the table above, % recovery is a ‘measure’ of the accuracy. This distribution of copper results in the “mean value” ranging between the extremes of poor accuracy (e.g. 50 or 150 % recovery) and good precision, good accuracy (say 100 ± 10% recovery) and poor precision; and poor accuracy and poor precision. With ‘n’ numbers of samples taken, it is noted that there are n+1 possible combinations, if the order is not important; e.g. 0.5 + 0.5 +1.5 is the same as 0.5 + 1.5 + 0.5 etc. If all the mean values from the n+1 distributions of a given ‘set’ are added together, they produce an overall mean value of 100% accuracy. For example, for n = 4, we have the “mean value” set of: 0.5 + 0.75 + 1.0 + 1.25 + 1.5 = 5.0 mg /5 = 1.0 mg, providing us with the correct overall concentration. This would of course assume an equal probability (weighting) for each distribution within the set being chosen! But as we will see a little later on (Chapter 8), this is not always the case. However, as demonstrated, as the

number of samples taken increases for the given distribution, the mean value approaches (but only approaches!) the correct value for the overall 'lot' analysed.

Problem 2

Before we leave this example, there are two further scenarios which we can consider. If the 32 sub-samples were to have a perfectly homogeneous distribution, so that a sub-sample with 0.5 mg of copper is always next to a sub-sample with 1.5 mg of copper in it, in all three dimensions, then what would be the effect of, i) taking one sub-sample at a time and sequentially, plotting the cumulative accuracy and precision as we go along, up to taking the full 32 sub-samples, and ii) taking twice as many sub-samples, two cubes each time, and looking at the accuracy and precision as we go along?

A Sampling Case Study

Consider the following case study where the mobility of the analyte of interest is noted and hence its distribution within the bulk or parent material, reflects this.

Case study 1) Water, water everywhere and every drop the same?

A stirred batch of drinking water was taken from a soft-drinks processing plant. The sample size was 1000 litres (approximately 1 metre x 1 metre x 1metre). The composition of a 1 mL sub-sample from this water = composition of the 1000 litre sample, in terms of, for example, it's Ca^{2+} ion content. Physical and physico-chemical properties of both sample and sub-sample, such as their temperature and pH would also be the same at the time the sub-sample was taken. While, during the production of this "batch" or "lot", the composition may vary slightly with time, the primary sample or "increment" (gross sample), will demonstrate that it is "homogeneous" and you wouldn't have to analyse the whole 1000 litres to obtain its composition, with regard for example once again, to its Ca^{2+} concentration. Here, the high mobility of the analyte of interest, together with its relatively large Ca^{2+} population (the concentration range is variable but ~ 5 to 50 mg/L is not uncommon for drinking waters; equivalent to 7.52×10^{19} to 7.52×10^{20} Ca^{2+} particles per litre) and its small particle size (the hydration sphere of a Ca^{2+} ion = ~ 2.3 to 2.7 x 10^{-10} m radius, [Byrd et al.; 2005]) is present in both the 1000 L gross sample (increment) and the sub-sample (1.000 mL test sample). This allows its distribution to be essentially identical despite the difference in sampling size being a factor of 10^6. In terms of the Ca^{2+} ion in the 1 mL test sample, its distribution using the above range of values would show:

The particle size = ~5.4 x 10^{-10} m hydrated diameter of Ca^{2+} ion

The particle volume = 82.4 x 10^{-30} m^3 = 82.4 x 10^{-24} mL

The number of particles in 1 mL of ~40 mg/L Ca^{2+} (~1 mM Ca^{2+}) = ~6.02 x 10^{17} ; a very, very large population number!

If evenly distributed, then each Ca^{2+} particle (82.4 x 10^{-24} mL) is within its own 'theoretical volume' of water of only ~1.66 x 10^{-18} mL. This means that, on average the distance between any two Ca^{2+} ion in solution will be 1.47 x 10^{-6} cm; that is only about 15 nm away from each other on average!

In other words, the highly uniform distribution within our stirred water sample would suggest that, at the macro-scale of cm or mL, the degree of heterogeneity → '0'

Now, in order to introduce an error of ~1% in the sampling population of the calcium, based upon its above distribution you would have to take a sample volume down around ~ 2 x 10^{-16} mL; equivalent to the volume of a nano-particle, some 70 nm in diameter. We don't generally take samples anywhere near this small in our macro world.

However, the same relative sampling error in population may be introduced by simply taking an incorrect sample volume of 0.99 mL instead of 1.00 mL; - just something to think about!

Problem 3

As an analyst working for a company that handles waste materials, you are called to the delivery office where a number of drums, nine in all, have been off-loaded. The relevant paperwork is to be signed-off but this is a new supplier of waste material and a new registered carrier. Checks are therefore required on the contents to see that all "matches up" with the documentation. The information provided states that the contents are waste "cutting oils" based upon emulsions of light mineral oil and water with a non-ionic surface active agent (detergent) added to stabilise the emulsion. These waste oils have been collected after being used as coolant and lubricant fluids for drilling and milling of metals at an engineering firm. The safety data sheet shows the material to be a light petroleum oil and water + detergent emulsion that is non-volatile, has low viscosity and is relatively safe to handle with advice on the handling requirements. What would be your sampling plan to verify the registered carrier's documentation on quantity and content of the nine x ~200 L drums?

Problem 4

A large tank, used for storing and delivering sugar water in a soft drinks production line at a packing facility is to be sampled*****. This is because there is a requirement to measure the sucrose levels in the tank prior to a major packing order run. The tank, which can be stirred holds nearly 800,000 L and has a diameter of just over 10 m across and is some 10 metres deep. Describe the sampling plan you would undertake to take a representative sample for measurement.

***** See in addition to chapter 2, the EPA document on “Tank Sampling” via the Environmental Protection Agency web site.

Problem 5 Boundary of a hot spot? to solve –

Using the equations shown in example 2.2 of Chapter 2 and the principles discussed in section 2.3.3, calculate the grid size needed and number of sampling points required for a sampling plan, covering a field site of 0.4 hectares, in order to identify the position of the boundary (within limits, ~ ± 1 metre) of a contaminant present as a ~circular hot spot of just on 10 metres in diameter, somewhere within the 0.4 hectare site with a confidence of 95%.

Problem 6: OTC medicine tablets produced from a batch

You are a quality control chemist for a pharmaceutical company. A drug, recently released from patent can now be sold over the counter (OTC) and you are heading up the analytical checks on the first few batches of produced material, to look at the packaged, processed drug. A full pilot batch of the drug is being made into pressed tablets and this limited production run is to check the tablets in their sealed ‘blister packs’. There are 8 processed tablets in each ‘pack’, and all packs are produced from the bulk batch. The packs are to be sampled from a moving conveyer which has three production streams (three packets in parallel coming off at any one time) and some 100,000 packets of tablets for the pilot production batch, which will take approximately 4.5 hours to complete the run.

What would be your sampling plan, if you were i) looking for damaged or improperly sealed packs from the entire run; and ii) required to determine that the contents of the tablets produced are within the specification of the labelling. For example, preliminary, smaller pilot runs prior to this larger, full scale production run all showed (from 10 analysed samples each) that the mean value of active drug in the tablets is ~ 510 mg but with a standard deviation of ~10.2 mg (a % RSD of 2.0).

A relatively detailed description of the processes involved is all that is required at this stage. The sampling errors and statistics for such a process will be covered later in Chapter 8 of the book and the on-line additions to Chapter 8.

Problem 7

Consider the list below, where five measurements are to be made:

The emission of $H_2S_{(g)}$ from a Landfill Waste Facility (TLV* and safety levels)

The levels of $SO_{2\,(g)}$ emitted from a gas-fired Power Station

The concentration of $NO_{2\,(g)}$ emitted from a diesel car engine

The emission of $HCl_{(g)}$ from an electrolysis-based precious metal plating bath

The concentration of $CO_{2\,(g)}$ in a carbonated (sparkling) water on a drinks packing line.

* Threshold Limiting Value

Now

i) What do these analytes all have in common (in terms of chemical and physical properties)?

And

ii) How might they be sampled for later estimation of their concentration?

Problem 8: Sampling "Battleships" Hot Spot

We saw in Chapter 2, Problem 2.1, that we could play an equivalent to the game of "Battleships" but based upon detecting the positions of a number of contaminants (called "Hot Spots") in a field using our "hotspots" calculations shown in example 2.2.

As we saw, there are 5 places in a field of given size that are contaminated but you don't know where your opponent has positioned them. Your task is to locate all 5 and the areas they each cover to discover their boundaries. You need help from your "Hot Spot" equations which allow you to construct the grids of the right size and arrangement to enable you, with confidence to make calculated guesses. Your opponent informs you if you have made a "hit" or not.

In this slightly extended version, the contaminated areas are not just circular but also different sized ellipses having variable axes. The size and shape of each "hot spot" is given to you, as before, by your opponent, together with the size of field.

An example: A field to be sampled is 0.64 Hectare (6,400 m^2 ; 80m x 80m if square) in size

and the 5 contaminated “hot spots” are of the following size and shape:

2 Circular hot spots, one of 5 m diameter, one of 15 m diameter

1 Elliptical hot spot of 10 m major axis and shape factor S = 0.6 (≡ 10 m x 6 m)

1 Elliptical hot spot of 20 m major axis and shape factor S = 0.5 (≡ 20 m x 10 m)

1 Elliptical hot spot of 30 m major axis and shape factor S = 0.7 (≡ 30 m x 21 m)

Now, from equation 2.1 (Chapter 2) the constant, 0.59, is based upon a ‘p’ value of 0.05 (95% CL) and a factor ‘S’ = 1.0 for circular shapes (see Gilbert, R.O., 1987 and Ferguson, C.C., 1992).

For these more challenging games, the shape factors of the elliptical contaminant areas are used. For reference these are shown below:

For elliptical hot spots the equation for the grid size, G is given by;

G = L / K where L = ½ length of the major axis of the ellipse.

Elliptical hot spot, 95% hit, S = 0.5 (minor/major axis = 0.5). Shape constant K = 0.9

Elliptical hot spot, 95% hit, S = 0.6 (minor/major axis = 0.6). Shape constant K = 0.8

Elliptical hot spot, 95% hit, S = 0.7 (minor/major axis = 0.7). Shape constant K = 0.72

Elliptical hot spot, 95% hit, S = 0.8 (minor/major axis = 0.8). Shape constant K = 0.67

Elliptical hot spot, 95% hit, S = 0.9 (minor/major axis = 0.9). Shape constant K = 0.62

Circular hot spot, 95% hit, S = 1.0 (minor/major axis = 1.0). Shape constant K = 0.59

[it is also noted that a circle is a special form of ellipse where the major and minor axes are equal!]
In effect you need to find by calculation the smallest grid size from the above (and therefore the maximum number of sampling points or grids) required for all your shapes. The one that produces the highest number of grids (from the ratio of smallest L and largest K values) becomes the controlling board layout, based upon that number.

Looking through the 5 shapes given in our example above, in terms of smallest L value, we have the 5 m circular hot spot where L = 5 / 2 = 2.5 m and K =0.59; therefore our grid size,

G = 2.5 / 0.59 = 4.2 m

And the number of sampling points, n (equation 2.2) would be:

$n = A / G^2$

n = 6400 / $(4.2)^2$

n = ~ 363 (nearest whole number)

You need to locate all 5 hot spots and the area they each cover to win the game.

The arrangement of grids will come from the root of the value for n.

$\sqrt{\mathbf{363}}$ = 19.0(5); approx. 19 x 19 grid arrangement.

As any arrangement would need to allow all shapes and sizes of hot spots to be identified. we might ensure we pick up all by using a whole number arrangement of 20 by 19.

This means the new, total number of sampling points is:

20 x 19 = 380. Our new grid size becomes:

G = $[6400 / 380]^{½}$; = 4.1 m

Your opponent then places the 5 shaped hot spots on the 20 by 19 grid arrangement , each grid 4.1 m apart (See below), without you knowing where, and your guesstimates begin using X and Y grid identifiers, as before.

As you would have supplied your opponent with their own size of field and their 5 shapes and sizes of ellipses for their challenge, the latter of which you can secretly place on the grid they have calculated, then your opponent can present their guesstimates, alternating with yours.

Remember that the shape of the hot spot will mean some grids are only partly occupied, but these must still be identified, in order to win.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				