Statistical consideration for sampling strategies – Part II

In the previous article on Sampling Strategies – Part I, we have discussed the simple random sampling method which provides an equal chance for a sample to be selected from a population. It has several advantages and disadvantages with certain restrictions.

Let's take a look at the other methods as described below.

2. **Stratified Random Sampling**

As the word 'stratified' suggests, the population being sampled is divided into segments (or strata) and a simple random sample is selected from each segment on a sequence of items. The Figure 1 shows how we can collect 5 random packet samples from a lot of 25 packaged products divided into 5 strata on the basis of sequence number.

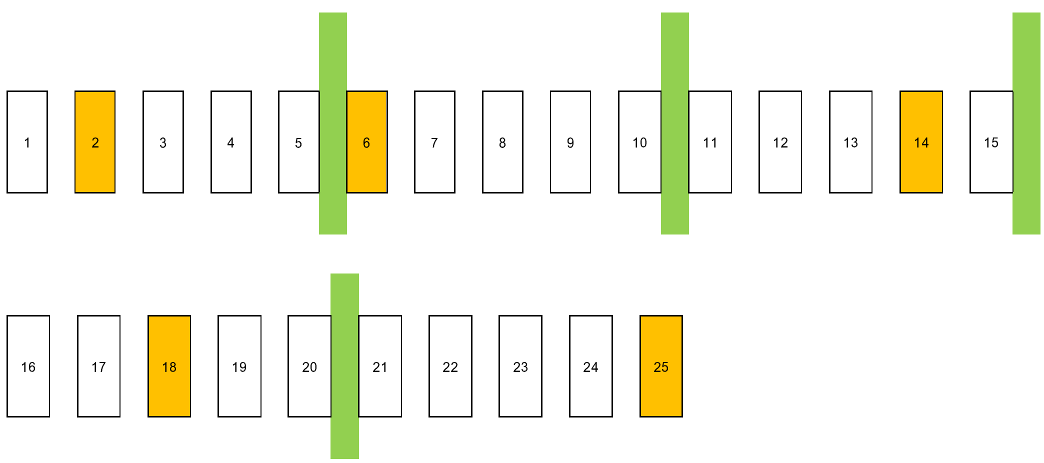

Figure 1:
Stratified random sampling for sequences. A stratified random sample of 5 items, shown in orange color, is formed by dividing the laboratory sample into 5 strata on the basis of sequence number, and selecting one sample at random from within each stratum. The greenish vertical lines show the divisions between strata. This limits the interval between selected items while still providing a random selection of items.

The number of items selected from each segment depends on the objective of the sampling exercise.

There are two particular important variants which are to be considered in such sampling exercise:

a. *To select equal number of items per stratum* : This is convenient and common when there is no information about the variance within each stratum. Composite samples may be formed within a stratum, but a composite of the whole does not show an unbiased estimate of the bulk sample composition unless all the strata are of equal size.

b. *To make proportional sampling* : .In this case, the number of items taken per stratum is proportional to the fraction of each stratum in the bulk population. This is ideal for forming a composite sample, as the expected mean for the composite is the same as the mean of the bulk. Sometimes we call it as a weighted average value. The variance is smaller than for simple random sampling.

Advantages

Where strata differ appreciably, stratified random sampling provides smaller sample variance than simple random sampling and can be planned to provide the minimum variance available.

Disadvantages

Stratified random sampling adds complexity in selecting the sample and calculating the mean value.

Statistical treatment

Statistical treatment for stratified random sampling is somewhat more complex than in simple random sampling.

Let us assume there are n_s strata, each containing N_i items with a total mass m_i, and that the measured mean, standard deviation and number of items tested in stratum i are $\bar{x}_i$, s_i and n_i , respectively.

In this case, the standard deviation s_i is assumed to be equivalent to the standard deviation of sampling, s_{sam}, that is, the standard deviation of observations for the stratum if the analytical variation is considered negligible, or, an estimate of the sampling variance alone if not.

The proportion P_i of each stratum in the bulk population is:

$$P_i = \frac{m_i}{\sum_{i=1,ns} m_i}$$

So, if we require the mean value $\bar{X}$ for the bulk, it is given by:

$$\overline{X} = \sum_{i-1,ns} P_i x_i$$

This mean value has a variance $s^2(\overline{X})$ given by:

$$s^2(\overline{X}) = \sum_{i=1,ns} \frac{P_i^2 s_i^{\ 2}(1-f_i)}{n_i}$$

where $f_i = n_i/N_i$.

The corresponding standard error, which can be used directly as the standard uncertainty from sampling, is given by:

$$s(\overline{X}) = \sqrt{s^2(\overline{X})}$$

The total mass of analyte required is calculated from the mean mass fraction $\overline{X}$ as: $m_{tot}\,\overline{X}$ with standard uncertainty from sampling equal to $m_{tot} s(\overline{X})$, m_{tot} being the total mass of the bulk material (population).

We illustrate the above statistics with an example below.

Example:

In a hypothetical case that police seized 35 bags from a suspected illegal drug seller, each containing 200g of white powder (totaled 7000g), later found to contain illicit drug A ith closely similar composition by IR spectroscopy.

In addition, 273 small packets each containing 5g of drug A with added ingredients (totaled 1365g) were found on the premises, again the IR spectroscopy confirmed that the smaller packets were closely similar in composition.

Now, let's make an estimate of the total amount of drug A.

With two distinct types of material and evidence of similarity within each, a stratified random sample is considered appropriate. So, in the absence of detailed variance information, 10 bags of each sample was taken and analyzed by HPLC, with the results shown in the upper part of Table1 below.

The subsequent calculations of the mean drug A concentration over all seized material are shown in the bottom part of the table (assuming that variation is predominantly from sampling), allowing the calculation of total drug A in the seizure and providing an uncertainty in the estimate.

Table 1: Results of sampling drug packages

	Large bags	Small packets	Whole consignment
Analysis results			
Stratum mass m_i (g)	7000	1365	
Total mass m_{tot} (g)			8365
Total No. of items N_i	35	273	
Number in test samples n_i	10	10	
Mean drug A concentration (g per 100g)	43.2	14.3	
Standard deviation of observations s_i (g per 100g)	3.5	2.7	
Calculations			
Proportion P_i	0.837	0.163	
$P_i \overline{x}_i$	36.2	2.3	
$f_i = n_i / N_i$	0.286	0.037	
Variance $p_i^2 s_i^2 (1 - f_i) / n_i$	0.613	0.019	
Mean drug concentration $\overline{X}$ (g per 100g)			(36.2+2.3) = 38.5
$s(\overline{X})$ (g per 100g)			Ö(0.613+0.019) = 0.795
Total drug A, $(m_{tot} \overline{X} / 100)$ (g)			3221
Std uncertainty $(m_{tots} (\overline{X}) / 100)$ (g) *			67

Note: * Due to sampling variation only