Random Sampling Strategies

As the new FDIS 17025 has emphasized the importance of sampling in addition to testing competence, let's now get familiar with some of the strategies available for sampling and see their various strengths and weaknesses.

1. Simple random sampling

In random sampling, each item in the population or sampling target for laboratory analysis has an equal chance of being selected through the use of a random number table or a random number generation from MS Excel® spreadsheet or some statistical software like R programming.

The usual method is to number the items sequentially, then use a table of random numbers or random number generated from software to select items randomly from the set.

For example, assume there is a cargo of 500 rice bags (20 kg each) to be surveyed for quality and we are requested to conduct a 5% sampling exercise (i.e. selecting 25 bags at random), chisel out say, about 200 gm rice from each bag and mix them into a composited sample for laboratory analysis.

To do the sampling, label each bag with a number and use Excel® random number generator to generate a random number to each and every bag. Use the =RANK() function to rank the bags in ascending order. Pick up the marked bags with rank order 1,20,40,60, …, 480, 500 from the warehouse as random samples and composite them accordingly. Figure 1 shows how to use the Excel spreadsheet functions for 500 bags (*viz.* A(1) to A(500)) selected at random and being ranked:

Figure 1: MS Excel® spreadsheet display of 500 items for ranking before arranging them in ascending order

	A	B	C
1	Item #	Random	Rank
2	A(1)	=RAND()	=RANK(A2,$A2:A501,1)
3	A(2)	=RAND()	=RANK(A3,$A2:A501,1)
4	A(3)	=RAND()	=RANK(A4,$A2:A501,1)
~~	~~	~~	~~
500	A(499)	=RAND()	=RANK(A500,$A2:A501,1)
501	A(500)	=RAND()	=RANK(A501,$A2:A501,1)

This random sampling method can be applied both in the field and in a laboratory setting for sub-sampling from a set of discrete items, such as ampoules or food packages.

For particulate materials such as resin or grainy beans, sub-samples from a bulk sample can be obtained by repeated cone-and-quartering process or rotatory sampling, with due care to avoid bias from size or density selection and to prevent loss of moisture or volatile oil in some sensitive volatile materials such as black and white pepper, due to heat generated during the sub-sampling process.

The aim of all these procedures is to generate sub-samples in which every member of the population (laboratory sample) has an equal chance of appearing in a sub-sample.

Advantages:

- Simple to implement;
- Selection is independent: selection of one unit has no influence on the selection of other units;
- Often sufficient for sub-sampling within a laboratory

Disadvantages:

- It can be subjective in preferred sampling at easily accessible location;
- Intuitive selection of either obviously darker or lighter color of the population
- Having very variable intervals between successive test items, making it a poor choice for monitoring items from a continuous sequence
- For non-homogeneous materials or for sample consisting of identifiable sub-groups with substantially different properties, the variance in this simple random sampling method is among the highest of the strategies described here. A printed circuit board sample, for example, contains many electronic parts with different amounts of ROHS (Restriction of Hazardous Substances) under the EC Directives.

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 2 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.

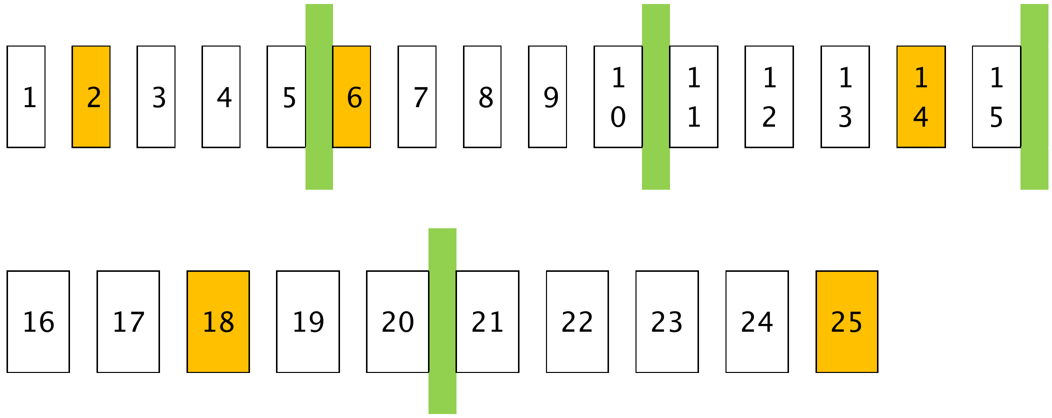

Figure 2:
Stratified random sampling for sequences. A stratified random sample of 5 items, shown in orange color, is formed by dividing the laboratory sample into 5 strata on the basis of sequence number, and selecting one sample at random from within each stratum. The greenish vertical lines show the 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.

3. Systematic sampling

As the term 'systematic' dictates, we have to select the first test sample at random and then pick up the subsequent samples at a fixed interval, such as every fifth item or at a fixed time during the streaming of items on a conveyor belt upon production.

For example, consider 1000 sequentially numbered items in a batch of a product of which 20 need to be selected for testing. What we have to do is to select the first at random from items marked 1-50, which every 50th item is chosen. If item number 9 was selected at random as the first test sample, the subsequent samples chosen would be 59, 109, 159, …, 959.

The main advantage for systematic sampling is due to its simplicity to implement over a period of time or a stream of material. In this type of sampling, the sample values are spread more evenly across the population, thus, many systematic samples are highly representative of the population from which they were selected. It also reduces the occurrence of long or very short intervals that may occur in simple random sampling.

However, systematic sampling can only be considered equivalent to random sampling if the variation in the material is reasonably independent of the item numbering or location. An extreme case can be that all items selected at intervals turn out to be of good quality whilst there might be bad ones in between! Thus, the precision of the test result obtained is mainly influenced by the 'gap' between the sampling points, and no valid estimate of sampling error can be calculated from a single sample taken.

Systematic samples are treated as if random, so the statistical treatment is as for simple random sampling as discussed above.

The next blog will discuss the basic random sampling theory