Experimental Design – Using randomization to reduce bias effects

As we know, most measurements represent straightforward application of a measuring device or method to a test sample. However, many experiments are intended to study the effects of the presence or absence of some specific treatments on measurement results, such as the effect of changing the experimental temperature or adjusting the sample size. We need therefore to plan a suitable design of experiment (DOE) with statistical consideration and replication in order to optimize the test method for performance.

Randomization is perhaps the most important tool to reduce any bias and uncontrollable factors, caused by both within-run and between-run effects in a design of experiment. It is strongly recommended in essentially all DOE to reduce the impact of unknown factors.

For example, we wish to conduct a series of experiments to assess whether two digestion methods are providing the same elemental analysis result. Let's label these methods A and B, and analyze a single reference material using 6 replicates of each digestion method. We label these sub-samples for preparation before instrumental measurements as follows:-

A(1), A(2), A(3), A(4), A(5), A(6), B(1), B(2), B(3), B(4), B(5), B(6)

Our practical experience tells us that instruments tend to drift steadily. If we were to run the experiments in accordance to the above order, we might obtain some systematic error in the whole experimentation. In order to control these nuisance effects, it is therefore best to randomly select these methods for experimentation.

How are we going to implement randomization?

Randomization depends on a source of random numbers to start with. Random numbers can be generated in a number of different ways:

1. Using random number tables

   Some statistics textbooks include random number tables. To use such tables, we start at any point in the table and read out numbers from that point onwards. Allocate the random numbers to the planned list of observations and on completion, re-order the observations in random number order.

2. Using established statistical or experimental design software

   All good statistical software includes a range of functions for generating random numbers, ranking and ordering values and often for generating randomly ordered samples from existing lists.

3. Using spreadsheet software

   Most spreadsheets include random number generators. In Excel, the relevant function is RAND( ). These are usually adequate for small experiments, although they do not perform as well as the best random number generators found in good statistical software.

   The Excel function RAND( ) can be used jointly with RANK( ) to create a randomization for the DOE as follows:

	A	B	C
1	Expt #	Random number	Rank
2	A(1)	=RAND()	=RANK(A2,$A2:A13,1)
3	A(2)	=RAND()	=RANK(A3,$A2:A13,1)
4	A(3)	=RAND()	=RANK(A4,$A2:A13,1)
~~	~~	~~	~~
11	B(4)	=RAND()	=RANK(A11,$A2:A13,1)
12	B(5)	=RAND()	=RANK(A12,$A2:A13,1)
13	B(6)	=RAND()	=RANK(A13,$A2:A13,1)

   By doing so, we got the following random numbers and their ranking:

Expt #	Random number	Rank
A(1)	0.1232	3
A(2)	0.4168	7
A(3)	0.7214	11
A(4)	0.5672	9
A(5)	0.0410	1
A(6)	0.2258	5
B(1)	0.5446	8
B(2)	0.1005	2
B(3)	0.3481	6
B(4)	0.6705	10
B(5)	0.9828	12
B(6)	0.1878	4

The experiments ware then carried out in accordance to the following ascending order:

Expt #	Random	Rank
A(5)	0.0410	1
B(2)	0.1005	2
A(1)	0.1232	3
B(6)	0.1878	4

A(6)	0.2258	5
B(3)	0.3481	6
A(2)	0.4168	7
B(1)	0.5446	8
A(4)	0.5672	9
B(4)	0.6705	10
A(3)	0.7214	11
B(5)	0.9828	12

However, it has to be stressed that randomization does have certain limitations. For example, the reduction in the uncontrollable effect is modest and depends on the number of replicates done.