DOE – Randomized block designs

In one of the previous blogs, we described Complete Randomized designs which are of one factor and are the simplest statistical designs of experiments. We used one-way ANOVA to analyze the data obtained.

Perhaps, it is worth mentioning that there are three basic principles of such statistical designs of experiments, namely control, randomization and repetition. Randomization is important in any experiment design to prevent any result bias, because the researcher cannot always be certain that every major influence on a response has been included in the experiment. It is also important to set a control (or so-called blank) in the treatment for comparison purposes.

For example in a hypothetical study of a medication drug efficacy in reducing fever involving some 1200 patients, it is good to subject also a group of patients (say 600 patients) taking placebo or a dummy drug without containing the active ingredient for treatment. Such dummy drug has similar color, smell, taste and even size as the medication drug under study. The subjects for the experiment are then assigned to taking either the placebo or the drug completely at random and the replicated measurements are then analyzed for their variances.

In some circumstances, however we need to have an improved understanding of the effects of factors or treatments, allowing more precision in the estimation of these effects, by removing some inherent effect that might contribute to the data variation which are of no direct interest in the study. In other words, we try to create blocks to reduce some unexplained variation of the sum of squares of error (SSE) of a completely Randomized designed experiment. We can group the experimental units or test runs into more "homogeneous" blocks where all levels of the main factor are equally represented. We can then randomly assign all factor combinations of interest in the experiment to the units or test sequence in a block, followed by the use of a separate randomization for each block.

In the above hypothetical drug experiment in reducing fever using 1200 patients, we can divide these subjects into two blocks based on gender and

assign each block to the treatment studies. We may then have 600 men and 600 ladies for the experiments which assign 300 of each sex to either the placebo or the treatment drug accordingly. Therefore, the randomized block deign has explicitly controlled the variable of gender in this instance.

In fact, blocks are another form of control which controls the variables that are used to form the blocks (we call it the blocking variables). In this example, the blocking variable is the gender. Two-way ANOVA is used to analyze the data outputs and is more efficient than the one-way ANOVA because the error variation is reduced. In this case, it is also assumed that:

1. The data are normally distributed

2. The populations considered have equal variances (homogeneity of variance)

3. The errors are independent because independent random samples are drawn

4. There is no interaction between blocks and treatments.

As with completely randomized designs, a simple model can be used to describe the general form of randomized block designs.

Let

- y_{ij} represent the data obtained from the experiment (the measured outcome or result) conducted on the jth replicate that receives the ith treatment;
- τ_i be the effect attributable to the ith treatment,
- β_j be the effect attributable to the jth block, and,
- ε denotes the residual error, unexplained by other factors in block j receiving treatment i.

then the statistical model for this kind of experiment is of the mathematical form:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

This model shows that the response y_{ij} is a simple linear function of the overall population mean value μ for all the data obtained, plus treatment and

block effects, and plus some residual error. Analysis of the data is then carried out by the two-way ANOVA procedure. This analysis of variance splits the error term of an equivalent single factor Complete Randomized design in block and error components.

Hence, the SST and $SSTr$ are the same as those in the completely randomized design whilst the error variance (SSE) has both a blocking effect (SSB) and a reduced error variance (SSE). Note that in a complete randomized design, there is no blocking effect (SSB) but only the error variance (SSE).

The following worked example should explain it more clearly.

A chemist carried out a series of chemical kinetic studies by following the first order reaction rates of a reactant X with three different types of catalysts per day over four days. Since the day-to-day differences may affect the reaction rate, each day was taken as a block and all three catalysts were tested each day independently in randomized orders. The reaction rates measured in moles min^{-1} were summarized as below:

Block (Day)	Catalyst A	Catalyst B	Catalyst C
1	0.30	0.33	0.34
2	0.28	0.29	0.33
3	0.31	0.29	0.32
4	0.30	0.34	0.35

For significance testing, we state the two hypotheses:

Ho : $\mu_1 = \mu_2 = \mu_3 = \mu$ (i.e. no differences amongst the reaction rates)
H_1 : Not all 3 reaction rate means (with catalyst factor) were equal

In this situation, we need to consider three independent estimates of the population variance σ^2:

- **Between-treatments (catalysts)** estimate of population variance
- **Between-blocks (days)** estimate of population variance
- **Within-blocks (error)** estimates of population variance

The general two-way ANOVA table used is as follows:

Sources of variations	Degree of freedom	Sum of squares (SS)	Mean square (Variance)	F
Treatment j	$p-1$	$SSTr$	$MSTr$	$MSTr/MSE$
Block i	$b-1$	SSB	MSB	MSB/MSE
Error	$n-p-b+1$	SSE	MSE	
Total	$n+1$	SST		

Given p= number of treatments, b = number of blocks, the formulae used are:

$SSTr = \sum_{j=1}^{p} b(\bar{x}_j - \bar{\bar{x}})^2$ with degree of freedom p–1

$SSB = \sum_{i=1}^{b} p(\bar{x}_i - \bar{\bar{x}})^2$ with degree of freedom b–1

$$SST = \sum_{i=1}^{n_j} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2$$

$$SSE = SST - SSTr - SSB$$

Hence, using the Excel Statistical Tools package for the *two-way ANOVA without replication,* we have:

ANOVA: Two-Way Without Replication

SUMMARY	Count	Sum	Average	Variance
Day 1	3	0.97	0.323	0.00043
Day 2	3	0.90	0.300	0.00070
Day 3	3	0.92	0.307	0.00023
Day 4	3	0.99	0.330	0.00070
Catalyst A	4	1.19	0.298	0.00016
Catalyst B	4	1.25	0.313	0.00069
Catalyst C	4	1.34	0.335	0.00017

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P*-value	*F* crit
Catalyst	0.00285	2	0.001425	6.66	0.03	5.14
Day	0.00177	3	0.000589	2.75	0.13	4.76
Error	0.00128	6	0.000214			
Total	0.0059	11				

Therefore $MSTr\,(Catalyst)$ = 0.00143, $MSB\,(Day)$ = 0.00059 and MSE = 0.00021.

Since in the catalyst factor, F=6.66>5.14, it is concluded that there were reaction rate differences amongst the use of catalysts (*treatments*), and since in the blocking, F=2.75< 4.76, then Ho is true, indicating that there was no significant difference amongst the reaction rate means done over the days.