Analysis of Variance and Design of Experiments-II

MODULE I

LECTURE - 1

INCOMPLETE BLOCK DESIGNS

Dr. Shalabh

Department of Mathematics & Statistics

Indian Institute of Technology Kanpur

If the number of treatments to be compared is large, then we need a large number of blocks to accommodate all the treatments. This requires more experimental material and so the cost of experimentation becomes high which may be in terms of money, labor, time, etc. The completely randomized design and randomized block design may not be suitable in such situations because they will require a large number of experimental units to accommodate all the treatments. In such situations when sufficient number of homogeneous experimental units are not available to accommodate all the treatments in a block, then incomplete block designs can be used. In incomplete block designs, each block receives only some of the selected treatments and not all the treatments. Sometimes it is possible that the available blocks can accommodate only a limited number of treatments due to several reasons.

For example, the goodness of a car is judged by different features like fuel efficiency, engine performance, body structure, etc. Each of these factor depends on many other factors, e.g., engine consists of many parts and the performance of every part combined together will result in the final performance of the engine. These factors can be treated as treatment effects. If all these factors are to be compared, then we need a large number of cars to design a complete experiment. This could be an expensive affair.

The incomplete block designs overcome such problems. It is possible to use much less number of cars with the set up of an incomplete block design and all the treatments need not to be assigned to all the cars. Rather some treatments will be implemented in some cars and remaining treatments in other cars. The efficiency of such designs is, in general, not less than the efficiency of a complete block design.

In an another example, consider a situation of destructive experiments, e.g., testing the life of television sets, LCD panels, etc. If there are large number of treatments to be compared, then we need a large number of television sets or LCD panels. The incomplete block designs can use lesser number of television sets or LCD panels to conduct the test of significance of treatment effects without losing, in general, the efficiency of design of experiment. This also results in the reduction of experimental cost.

Similarly, in any experiment involving the animals like as biological experiments, one would always like to sacrifice less animals. Moreover, the government guidelines also restrict the experimenter to use a smaller number of animals. In such cases, either the number of treatments to be compared can be reduced depending upon the number of animals in each block or to reduce the block size.

In such cases when the number of treatments to be compared is larger than the number of animals in each block, then the block size is reduced and the setup of incomplete block designs can be used. This will result in the lower cost of experimentation. The incomplete block designs need less number of observations in a block than the observations in a complete block design to conduct the test of hypothesis without losing the efficiency of design of experiment, in general.

Complete and incomplete block designs

The designs in which every block receives all the treatments are called the **complete block designs**.

The designs in which every block does not receive all the treatments but only some of the treatments are **called incomplete block design.**

The block size is smaller than the total number of treatments to be compared in the ncomplete block designs.

There are three types of analysis in the incomplete block designs

- intrablock analysis,
- interblock analysis and
- recovery of interblock information.

Intrablock analysis

In intrablock analysis, the treatment effects are estimated after eliminating the block effects and then the analysis and the test of significance of treatment effects are conducted further. If the blocking factor is not marked, then the intrablock analysis is sufficient enough to provide reliable, correct and valid statistical inferences.

Interblock analysis

There is a possibility that the blocking factor is important and the block totals may carry some important information about the treatment effects. In such situations, one would like to utilize the information on block effects (instead of removing it as in the intrablock analysis) in estimating the treatment effects to conduct the analysis of design. This is achieved through the interblock analysis of an incomplete block design by considering the block effects to be random.

Recovery of interblock information

When both the intrablock and the interblock analysis have been conducted, then the two estimates of treatment effects are available from each of the analysis. A natural question then arises -- Is it possible to pool these two estimates together and obtain an improved estimator of the treatment effects to use it for the construction of test statistic for testing of hypothesis? Since such an estimator comprises of more information to estimate the treatment effects, so this is naturally expected to provide better statistical inferences. This is achieved by combining the intrablock and interblock analysis together through the recovery of interblock information.

Intrablock analysis of incomplete block design

We start here with the usual approach involving the summations over different subscripts of y's. Then gradually, we will switch to matrix based approach so that the reader can compare both the approaches. They can also learn the one –to-one relationships between the two approaches for better understanding.

Notations and normal equations

Let

> ➢ v treatments have to be compared.

> ➢ b blocks are available.

> ➢ k_i : Number of plots in ith block ($i = 1,2,…,b$).

> ➢ r_j : Number of plots receiving jth treatment ($j = 1,2….,v$).

> ➢ n : Total number of plots.

> $n = r_1 + r_2 + ... + r_v = k_1 + k_2 + ... + k_b.$

> ➢ Each treatment may occur more than once in each block

> or

> may not occur at all.

> ➢ n_{ij} : Number of times the jth treatment occurs in ith block.

For example, $n_{ij} = 1$ or 0 for all i, j means that no treatment occurs more than once in a block and a treatment may not occur in some blocks at all. Similarly, $n_{ij} = 1$ means that j^{th} treatment occurs in i^{th} block and $n_{ij} = 0$ means that j^{th} treatment does not occurs in i^{th} block.

$$\sum_{j=1}^{v} n_{ij} = k_i \qquad i = 1,...,b$$

$$\sum_{i} n_{ij} = r_j \qquad j = 1,...,v$$

$$n = \sum_{i} \sum_{j} n_{ij}$$

Model

Let y_{ijm} denotes the response (yield) from the mth replicate of jth treatment in i^{th} block and

$$y_{ijm} = \beta_i + \tau_j + \varepsilon_{ijm} \quad i = 1, 2, ..., b, \ j = 1, 2, .., v, \ m = 1, 2, ..., n_i$$

[**Note**: We are not considering here the general mean effect in this model for better understanding of the issues in the estimation of parameters. Later, we will consider it in the analysis.]

Following notations are used in further description.

Block totals : $B_1, B_2, ..., B_b$ where $B_i = \sum_j \sum_m y_{ijm}$

Treatment totals: $V_1, V_2, ..., V_v$ where $V_j = \sum_i \sum_m y_{ijm}$

Grand total : $Y = \sum_i \sum_j \sum_m y_{ijm}$

Generally, a design is denoted by $D(v, b, r, k, n)$ where *v, b, r, k* and *n* are the parameters of the design.

Normal equations

Minimizing $S = \sum_i \sum_j \sum_m \varepsilon_{ijm}^2$ with respect to β_i and τ_j, we obtain the least squares estimators of the parameters as follows:

$$S = \sum_i \sum_j \sum_m (y_{ijm} - \beta_i - \tau_j)^2$$

$$\frac{\partial S}{\partial \beta_i} = 0$$

$$\Rightarrow \sum_j \sum_m (y_{ijm} - \beta_i - \tau_j) = 0$$

or

$$B_i - \beta_i \sum_j \sum_m 1 - \sum_j \tau_j \sum_m 1 = 0 \qquad (1)$$

or

$$B_i = \beta_i k_i + n_{i1}\tau_1 + n_{i2}\tau_2 + ... + n_{iv}\tau_v, \ \ i = 1,...,b$$

$$B_i = \beta_i k_i + \sum_j \tau_j n_{ij} \qquad [b \text{ equations}]$$

$$\frac{\partial S}{\partial \tau_i} = 0$$

$$\Rightarrow \sum_i \sum_m (y_{ijm} - \beta_i - \tau_j) = 0$$

or $$\sum_i \sum_m y_{ijm} - \sum_i \beta_i \sum_m 1 - \tau_j \sum_i \sum_m 1 = 0$$

$$V_j - \sum_i \beta_i n_{ij} - \tau_j \sum_i n_{ij} = 0 \qquad (2)$$

or $$V_j = n_{ij}\beta_1 + n_{2j}\beta_2 + ... + n_{bj}\beta_b + r_j \tau_j, \ \ j = 1,2,...,v$$

or $$V_j = \sum_i \beta_i n_{ij} + r_j \tau_j \qquad [v \text{ equations}]$$

Equations (1) and (2) constitute ($b + v$) equations.

Note that

$$\sum_i \text{equation (1)} = \sum_j \text{equation (2)}$$

$$\sum_i B_i = \sum_j V_j$$

$$\sum_i \left(\sum_j \sum_m y_{ijm} \right) = \sum_j \left(\sum_i \sum_m y_{ijm} \right).$$

Thus there are at most ($b + v$ - 1) degrees of freedom for estimates. So the estimates of only ($b + v$ - 1) parameters can be obtained out of all ($b + v$) parameters.

[**Note**: We will see later that degrees of freedom may be less than or equal to ($b + v$ - 1) in special cases. Also note that we have not assumed any side conditions like $\sum_i \alpha_i = \sum_j \beta_j = 0$ as in the case of complete block designs.]