Design of Experiments

6.1. INTRODUCTION

Design of an experiment is essential to draw meaningful and valid conclusions for specified objectives under a statistical study. Design of any experiment involves planning, obtaining relevant information regarding the statistical hypothesis under study and making statistical analysis of the data. Planned and well-designed experiments give us more accurate and fruitful results.

For example, an experiment is conducted to examine the effect of three different drugs administered to five patients. To conduct this experiment efficiently, first the experimentor has to plan the major concerns like selection of suitable and appropriate patients which are relevant to the study, under which climatic conditions the drug has to be administered, how much quantity of drug is to be applied, the better way of taking the measurements etc. Proper care has to be taken to formulate statistical hypothesis and to obtain relevant information to study it. Eventually an efficient statistical analysis has to be made to obtain valid inferences. This type of planning is called design of experiments.

6.2. DEFINITION

The design of experiment may be defined as the logical construction of the experiment in which the degree of uncertainty with which the inference is drawn may be well defined.

6.3. TERMINOLOGY IN EXPERIMENTAL DESIGN

Experiment

Experiment means getting an answer to a specified problem under consideration.

There are two kinds of experiments (1) Absolute experiments and (2) Comparative experiments.

(1) Absolute Experiments

Absolute experiments are designed to determine the absolute value or precise value of some characteristic of the study.

Examples :

1. Determining the average marks in a subject in a college.
2. Finding the correlation coefficient between income and expenditure of a group of selected families.

(2) Comparative Experiments

Comparative experiments are designed to compare the effect of two or more objectives of some population characteristics.

Examples :

1. A comparative study of three fertilizers on a crop in different plots of a field.
2. A comparative analysis between two different drugs made by a company.
3. A comparative study of different kinds of cultivation processes.

Treatment

Treatments are various objects of comparison applied on one or more experimental units in a comparative experiment.

Examples :

1. If five fertilizers are applied on various plots in the agricultural land, then fertilizers are treatments.
2. If three cultivation processes are applied on various plots in the field of agriculture, then cultivation processes are called treatments.
3. If three kinds of drugs are applied to five patients to make a comparative study, then drugs are known as treatments.

Experimental Unit

The smallest division of the experimental material to which we apply the treatments and on which we make observations on the variable under study is termed as an experimental unit.

Examples :

1. In the agricultural field experimentation, the smallest division or part is a plot. *i.e.*, plot is an experimental unit.
2. In a medical experiment, the smallest division is patient, hence patient is an experimental unit.
3. The experiment relates with cows, cow is treated as an experimental unit.

Blocks

If the whole experimental material is divided into relatively homogeneous sub-groups or strata, then these strata or sub-groups are known as blocks.

Example : In agricultural experiments, we divided the whole experimental material into relatively homogeneous sub-groups. These sub-groups are called blocks.

Yield

The measurement of the variable under study on different experimental units is known as yield.

Examples :

1. If an agricultural experiment relates to cultivating paddy, for example 70 kg of paddy is obtained from a plot (experimental unit), then 70 kg of paddy is treated as yield from the respective experimental unit *i.e.*, plot.

Experimental Error

If we allocate the same treatments for all the experimental units of the homogeneous experimental material, but the yields still vary from experimental unit to unit. The variation (apart from the variation due to known factors). which does not follow any systematic pattern and due to random factors beyond human control is known as experimental error or error. It is also called Noise.

Example : If we consider an agricultural experiment on which plots (experimental units) are homogeneous of equal shape or size in the field (experimental material) even if the same treatments are applied to all the plots, still the yield vary from plot to plot due to the differences in soil fertility. Such variation from plot to plot due to random factors (soil fertility) is called an experimental error.

6.4. IMPORTANCE OF DESIGN OF EXPERIMENTS

1. A proper design of an experiment leads to accurate results, this helps in achieving the objectives of the experiment.
2. Design of experiments is more useful in reducing the effects of errors due to various sources.
3. The design of an experiment increases the precision of the experiment.
4. The principles of design of experiments support in increasing the efficiency of the experiment.
5. The principle of design of experiments are useful in eliminating the personal bias in any form.

6.5. APPLICATIONS OF DESIGN OF EXPERIMENTS

1. Design of experiments is extensively used in agriculture, green house studies, etc.
2. The design of experiments is applied to study the laboratory techniques, medical experiments, pharmaceutical experiments, etc.
3. Design of experiments is used in the field of economics and industry.

6.6. PRINCIPLES OF AN EXPERIMENTAL DESIGN

According to R.A. Fisher, the basic principles of an experimental design are :

1. Replication

2. Randomisation
3. Local control.

6.7. REPLICATION

Replication means repetition of treatments more than once under investigation, *i.e.*, executing the experiment more than once.

(or)

Replication is a process so repeating the same treatment on various experimental units under similar conditions.

Replication is applied to average out the influence of the chance factors on different experimental units. Thus, the replication results in more reliable estimates. If a treatment is replicated 'r' times to the experimental units, then the standard error of the mean is $\frac{\sigma}{\sqrt{r}}$.

Advantages

1. Replication is useful to reduce experimental error and thus it enables us to obtain more precise estimates of the treatment effects.
2. The large number of replications reduce the standard error and it results in increasing the precision.
3. One of the most important purpose of replication is to provide an estimate of the experimental error.

Disadvantages

1. Increasing the number of replications is not possible since it involves more cost.
2. Replication of treatments may be sometimes subject to bias.
3. Large number of replications may lead to scarcity of the resources.

Notes :

1. It is desirable to have as homogeneous as possible within each replication, but homogeneity is not required between the replications.
2. The adequate number of replications to various treatments depends on the variability of the experimental material.
3. The number of replications considered in the experimental should provide at least 12 degrees of freedom for the error. Hence usually we should consider number of replications not less than 4.

6.8. RANDOMISATION

A process of allocating the treatments to various experimental units in such a way that each and every experimental unit has an equal chance of receiving any of the treatments is called randomisation.

The main objective of randomisation is validity of the statistical test of significance. All the statistical tests (*e.g.*, t-test, F-test, normal test etc.) are based on the fundamental assumption that the observations and drawn randomly *i.e.*, independent.

Advantages

1. Randomisation eliminates any kind of human bias.
2. It provides logical basis for conducting various statistical tests of significance.
3. It ensures the experiment free from any systematic effects of environment.

Disadvantages

1. Randomisation is difficult to apply of large number of experimental units.
2. Randomisation process can be effectively applied by statisticians only.

Note : Randomisation without replication is not sufficient to apply the various statistical tests of significance.

6.9. LOCAL CONTROL

If the experimental material is heterogeneous and different treatments are applied randomly over the entire material. If the experimental error is increased due to some uncontrolled factors even the treatments are applied randomly in sufficient number of times. Then it is desirable to reduce the experimental error without increasing the number of replications and without interferring the principle of randomness. This is achieved by principle of local control.

Definition

The process of reducing the experimental error by dividing the relatively heterogeneous experimental material into homogeneous blocks or subgroups is known as local control.

For example, in agricultural experimentation, the heterogeneity is observed in various experimental units, *i.e.* plots. The experimental error is increased due to fertility gradient of soil (or fluctuations in the field) even if different fertilizers (treatments) have been applied randomly with sufficient replications (number of times). In addition to the principles of replication and randomization, the experimental error can further be reduced by dividing or grouping the agricultural experimental field into relatively homogeneous blocks than those widely spread, *i.e.*, the entire field is divided into homogeneous blocks according to the fertility gradient of soil such that the variations with in each block is minimum and between the blocks is maximum. And then the treatments are allocated randomly with in each block.

Advantages

1. Local control reduces the experimental error.
2. Local control ensures the experimental design is more efficient.
3. By reducing the experimental error to a great extent, we can detect even small differences between the treatments.

Disadvantages

1. If the experimental material is homogeneous, then dividing into blocks is not necessary *i.e.* application of local control principle is not much useful.
2. In general, soil fertility does not follow any systematic pattern, hence dividing into homogeneous blocks is a difficult task.

◆◇◆

Completely Randomised Design (CRD)

7.1. INTRODUCTION

The completely randomised design (CRD) is the simplest of all the designs and it is based on the principles of replication and randomization. This design is suitable for which the experimental material is more homogeneous. In this design treatments are allocated randomly to all the experimental units over the entire experimental material. Let us suppose that we have k treatments, ith treatment is being replicated n_i times, $i = 1, 2, \ldots, k$ such that the total experimental material is $\sum_{i=1}^{k} n_i = N$. If all the treatments are replicated equal number of times, then total experimental material is $N = nk$.

7.2. DEFINITION

If the treatments are allocated to all the experimental units randomly over the entire experimental material by using the principles of randomization and replication is called completely randomised design (CRD).

7.3. ADVANTAGES, DISADVANTAGES AND APPLICATIONS

Advantages

1. CRD is more flexible since any number of treatments can be used and various treatments can be used unequal number of times without complicating the statistical analysis.
2. Statistical analysis remains simple if some or all the observations for any treatment are rejected or lost or missing for some random accidental reasons. Moreover the loss of information is smaller than any other design.

3. CRD results in the maximum use of the experimental units since all the experimental material can be used.
4. CRD provides the maximum number of degrees of freedom for the estimation of error variance, which increases the precision of the experiment.

Disadvantages

1. CRD can be used only for homogeneous experimental material. It is not possible always. Hence CRD is seldom used in field experimentation.
2. Since the randomization is not restricted in any direction, the variations among the experimental units is included in the error variance which make the design less efficient.
3. Since the principle of local control is not used, it results in increase of experimental error.

Applications

CRD is most useful in laboratory techniques and methodological studies *e.g.*, in physics, chemistry, etc. or in biological, chemical experiments or in green house studies, etc.

7.4. STATISTICAL ANALYSIS OF CRD

The statistical analysis of CRD is similar to the ANOVA for one way classified data.

The linear mathematical model becomes

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \begin{matrix} i = 1, 2, \ldots, k \\ j = 1, 2, \ldots, n_i \end{matrix}$$

where

y_{ij} is the yield from the jth unit by receiving the ith treatment.

μ is general mean effect

α_i is effect due to ith treatment

ε_{ij} is error effect due to chance and $\varepsilon_{ij} \sim iid\ \mathrm{N}(0, \sigma_e^2)$

$\mathrm{N} = \sum_{i=1}^{k} n_i$ is total number of experimental units.

k is number of treatments

n_i is number of replications of ith treatment.

$\mathrm{G} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$ is grand total.

$\mathrm{T}_{i.} = \sum_{j=1}^{n_i} y_{ij}$ is total yield of the units receiving the ith treatment

$\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ is mean yield due to ith treatment.

$\bar{y}_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$ is overall mean.

Null Hypothesis

$H_T : \alpha_1 = \alpha_2 = \ldots\ldots = \alpha_k$ *i.e.* all the treatment effects are same.

(or) all the treatment effects do not differ significantly.

Test a Statistic (or) Variance Ratio

Under the null hypothesis H_T, F statistic becomes

$$F_T = \frac{S_T^2}{S_E^2} \sim F_{(k-1,\, N-k)}$$

where $S_T^2 = \frac{SST}{k-1}$ is MSS due to treatments

$S_E^2 = \frac{SSE}{N-k}$ is MSS due to error

Conclusion

If calculated value of F is greater than the significant value (tabulated value) at specified level of significance *i.e.*, if $F > F_{(k-1,\, N-k)}$ then H_T may be rejected. Otherwise H_T may be accepted.

ANOVA Table for CRD

Source of variation	*Sum of squares*	*d.f.*	*Mean sum of squares*	*F-ratio (Variance ratio)*
Treatment	SST	$k-1$	$S_T^2 = \frac{SS}{k-1}$	$F = \frac{S_T^2}{S_E^2} \sim F_{(k-1,\, N-k)}$
Error	SSE	$N-k$	$S_E^2 = \frac{SSE}{N-k}$	
Total	TSS	$N-1$		

Randomised Block Design (RBD)

8.1. DEFINITION

A method of dividing the heterogeneous experimental material into relatively homogeneous subgroups or blocks or strata and the treatments are applied randomly to relatively homogeneous experimental units within each block and replicated over all the blocks is known as Randomised Block Design (RBD).

For example, in agricultural experimentation, the experimental area *i.e.* field is not homogeneous and the fertility gradient is only in one direction, then a simple method of controlling the variability of the experimental material consists dividing the heterogeneous field into homogeneous subgroups or blocks perpendicular to the direction of the fertility gradient. Then apply the treatments randomly to relatively homogeneous experimental units within each block and replicated over all blocks.

8.2. ADVANTAGES, DISADVANTAGES AND APPLICATIONS

Advantages

1. RBD provides more accurate results than CRD since the experimental material is divided into blocks this results in decreasing error variance.
2. In RBD, no restrictions are placed on the number of treatments and on the number of replications. But atleast two replications are required to test the significance of treatments or blocks.
3. Statistical analysis of RBD is simple and rapid.
4. Statistical analysis is simple even if the data is missed for any number of treatments.

5. The number of replications provide enough degrees of freedom for error sum of squares.

Disadvantages

1. RBD is not suitable for large number of treatments. Since block may contain more variability.
2. The error may be increased if the interaction between the treatments and blocks exist.
3. In every type of experiment blocking or grouping is not possible.

Applications

RBD is most useful in the field experiments like agriculture, environmental studies etc.

8.3. LAYOUT OF RBD

The layout of RBD is good enough to explain through an example. In agricultural experiment, if we consider six fertilizers (treatments) are applied (replicated) four times. Then we divide the whole experimental area (field) is divided into four relatively homogeneous blocks and each block consists of six units, six fertilizers are to be applied randomly to all six units within each block so that a treatment should be replicated four times. Let the 6 treatments are A, B, C, D, E, F, then layout of RBD can be explained below :

Blocks	***Treatments***					
I	B	C	D	A	E	F
II	C	F	D	E	A	B
III	F	A	E	B	D	C
IV	C	E	A	D	B	F

8.4. STATISTICAL ANALYSIS OF RBD

The statistical analysis of RBD is similar to the ANOVA for two way classified data. The linear mathematical model becomes

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$
$$i = 1, 2, \ldots, k; j = 1, 2, \ldots, h$$

where

y_{ij} is the yield from jth block by receiving ith treatment.

μ is general mean effect

α_i is effect due to ith treatment

β_j is effect due to jth block

ε_{ij} is error effect due to chance and $\varepsilon_{ij} \sim iid\ N(0, \sigma_e^2)$

$N = kh$ is the total number of experimental units.

k = number of treatments

h = number of blocks or replicates.

$$G = \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij} \text{ is grand total}$$

$$T_{i.} = \sum_{j=1}^{h} y_{ij}, \quad T_{.j} = \sum_{i=1}^{k} y_{ij}$$

$$\bar{y}_{i.} = \frac{1}{h} \sum_{j=1}^{h} y_{ij}, \quad \bar{y}_{.j} = \frac{1}{k} \sum_{i=1}^{k} y_{ij}$$

$$\bar{y}_{..} = \frac{1}{kh} \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij}$$

Null Hypothesis

1. For treatment

   $H_T : \alpha_1 = \alpha_2 = \ldots = \alpha_k$

   i.e., all treatments effects do not differ significantly.

2. For blocks

   $H_B : \beta_1 = \beta_2 = \ldots = \beta_h$

   i.e., all block effects do not differ significantly.

Test a Statistic (or) Variance Ratio

1. Under the null hypothesis H_T, Variance ratio becomes

$$F_T = \frac{S_T^2}{S_E^2} \sim (k-1, (k-1)(h-1))$$

where S_T^2 = MSS due to treatments

$$= \frac{SST}{k-1}$$

$$S_E^2 = \text{MSS due to error} = \frac{SSE}{(k-1)(h-1)}$$

2. Under the null hypothesis H_B, variance ratio becomes

$$F_B = \frac{S_B^2}{S_E^2} \sim F_{(h-1, (k-1)(h-1))}$$

where S_B^2 = MSS due to blocks $= \frac{SSB}{h-1}$.

Conclusion

If calculated value of F (*i.e.*, F_T or F_B) is greater than the significant value (tabulated value) at specified level of significance *i.e.*, if $F_T > F_{(k-1, (k-1)(h-1))}$, then we may reject H_T. Otherwise we may accept H_T. If $F_B > F_{(h-1, (k-1)(h-1))}$, then we may reject H_B. Otherwise we may accept H_B.

ANOVA Table for RBD

Source of variation	*Sum of squares*	*d.f.*	*Mean sum of squares*	*F-ratio (Variance ratio)*
Treatment	SST	$k-1$	$S_T^2 = \frac{SST}{k-1}$	$F_T = \frac{S_T^2}{S_E^2} \sim F_{(k-1, (k-1)(h-1))}$
Blocks	SSB	$h-1$	$S_B^2 = \frac{SSB}{h-1}$	$F_B = \frac{S_B^2}{S_E^2} \sim F_{(h-1, (k-1)(h-1))}$
Error	SSE	$(k-1)(h-1)$	$S_E^2 = \frac{SSE}{(k-1)(h-1)}$	
Total	TSS	$kh-1 = N-1$		

Latin Square Design (LSD)

9.1. DEFINITION

The entire heterogeneous experimental material is divided into relatively homogeneous blocks with respect to two sources of variation rows and columns, treatments are now allocated to all the experimental units in rows and columns in such a way that every treatment occurs once and only once in each row and in each column, such a design is called Latin Square design.

9.2. COMPARISON WITH RBD

Usually in RBD the whole experimental area is divided into relatively homogeneous groups or blocks and then treatments are allocated at random to the experimental units within eacn block *i.e.* randomisation principle is applied within each block *i.e.* randomisation is restricted in one direction (*i.e.* in blocks). But, generally in field experimentation (like agricultural experiments), the experimental area exhibits fertility in strips, for example high, medium, low fertility occurs in alternative strips of the field. RBD will be effective if the blocks are divided into parallel to the strips and RBD is extremely inefficient if the blocks are across the strips. Initially the pattern of fertility is seldom known. A very useful method of eliminating variations in the fertility gradient is controlling the variability in two perpendicular directions. This can be achieved with Latin square design.

9.3. ADVANTAGES AND DISADVANTAGES

Advantages of LSD

1. LSD controls more variation than CRD and RBD. Since error sum of squares is reduced by two way elimination of variation.
2. The statistical analysis is simple even with missing data.
3. More than one factor can be studied simultaneously and with fewer trials than more complicated designs.

Disadvantages of LSD

1. The number of treatments is equal to the number of replications in LSD. This limitation restricted the use of LSD. Hence, in general, LSD is suitable for the number of treatments between 5 and 10. The design is not suitable and impracticable for more than 10 treatments.
2. If several units are missing in LSD, the statistical analysis is more difficult.
3. RBD can be performed equally well on a square and rectangular field also, but LSD can only be performed on a square field.
4. The fundamental assumption that there is no interaction between different factors may not be true in general.

9.4. LAYOUT OF LSD

In LSD, the number of treatments and number of replications (blocks) are equal *i.e.* number of rows is equal to number of columns. If we consider m treatments, then there will be $m \times m = m^2$ experimental units. The whole experimental material is divided into m^2 experimental units arranged in a square so that each row and each column consists m experimental units. Then m treatments are allocated at random to these rows and columns in such a way that each and every treatment occurs once and only once in each row and in each column. This layout is called $m \times m$ LSD.

For example, if there are five treatments, then 5×5 LSD layout can be explained in the following table.

B	E	A	C	D
E	A	D	B	C
C	D	B	E	A
A	B	C	D	E
D	C	E	A	B

9.5. STATISTICAL ANALYSIS OF LSD

Let us suppose that y_{ijk} $(i, j, k = 1, 2, \ldots, m)$ be the yield from the experimental unit in the ith row, jth column and by receiving the kth treatment. The triplet (i, j, k) assumes m^2 experimental units in LSD. Let us denote the set of m^2 values by S. Usually we write $(i, j, k) \in S$. The linear mathematical model of LSD becomes.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \; (i, j, k) \in S.$$

where μ = general mean effect

α_i = effect due to the ith row

β_j = effect due to the jth column

γ_k = effect due to the kth treatment.

ε_{ijk} = error effect due to chance and $\varepsilon_{ijk} \sim iid\ N(0, \sigma_e^2)$

Consider

$G = y_{...}$ = Total of all the m^2 observations

$R_i = y_{i..}$ = Total of m observations in the ith row.

$C_j = y_{.j.}$ = Total of m observations in the jth column.

$T_k = y_{..k}$ = Total of m observations from the kth treatment.

Null Hypothesis

1. For Row effects :

   $H_R : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$

   i.e., row effects do not differ significantly.

2. For column effects :

   $H_C : \beta_1 = \beta_2 = \ldots\ldots = \beta_m = 0$

   i.e., column effects do not differ significantly.

3. For treatment effects :

   $H_T : \gamma_1 = \gamma_2 = \ldots\ldots = \gamma_k = 0$

   i.e., treatment effects do not differ significantly.

Test a Statistic (or) Variance Ratio

1. Under the null hypothesis H_R, variance ratio is

$$F_R = \frac{S_R^2}{S_E^2} \sim F_{((m-1),\,(m-1)(m-2))}$$

where S_R^2 = MSS due to Rows = $\dfrac{SSR}{m-1}$

$$S_E^2 = \text{MSS due to error} = \frac{SSE}{(m-1)(m-2)}$$

2. Under the null hypothesis H_C, variance ratio is

$$F_C = \frac{S_C^2}{S_E^2} \sim F_{((m-1),\,(m-1)(m-2))}$$

where S_C^2 = MSS due to columns = $\dfrac{SSC}{m-1}$

3. Under the null hypothesis H_T, variance ratio is

$$F_T = \frac{S_T^2}{S_E^2} \sim F_{((m-1),\,(m-1)(m-2))}$$

where S_T^2 = MSS due to treatments = $\dfrac{SST}{M-1}$

Conclusion

If calculated value of F (*i.e.*, F_R or F_C or F_T) is grater than the significant value (or tabulated value) at specified level of significance *i.e.*,

if $F_R > F_{(m-1,\,(m-1)(m-2))}$, then we may reject H_R, otherwise we may accept H_R.

if $F_C > F_{((m-1),\,(m-1)(m-2))}$, then we may reject H_C, otherwise we may accept H_C.

if $F_T > F_{((m-1)(m-1)(m-2))}$, then we may reject H_T, otherwise we may accept H_T.

ANOVA Table for LSD

Source of variation	*Sum of squares*	*d.f.*	*Mean sum of squares*	*F-ratio (Variance ratio)*
Rows	SSR	$m-1$	$S_R^2 = \dfrac{SSR}{m-1}$	$F_R = \dfrac{S_R^2}{S_E^2} \sim F_{(m-1,\,(m-1)(m-2)}$
Columns	SSC	$m-1$	$S_C^2 = \dfrac{SSC}{m-1}$	$F_C = \dfrac{S_C^2}{S_E^2} \sim F_{(m-1,\,(m-1)(m-2)}$
Treatments	SST	$m-1$	$S_T^2 = \dfrac{SST}{m-1}$	$F_T = \dfrac{S_T^2}{S_E^2} \sim F_{(m-1,\,(m-1)(m-2)}$
Error	SSE	$(m-1)(m-2)$	$S_E^2 = \dfrac{SSE}{(m-1)(m-2)}$	
Total	TSS	m^2-1		

8.8. MISSING PLOT TECHNIQUE IN RBD

If any observation is missed due to some unknown reasons, in RBD, we merely carry out the statistical analysis. The missing value is estimated by minimising the error sum of squares. This technique is called missing plot technique in RBD.

Let the observation $y_{ij} = x$ (say) is missing in RBD in the jth block by receiving the ith treatment. The two way table can be explained below as follows :

Treatments	***Blocks***						***Means***
	1	2	...	j	...	h	
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1h}	$T_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2h}	$T_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
i	y_{i1}	y_{i2}	...	$y_{ij} = x$	...	y_{ih}	$T_{i.} = T'_{i.} + x$
⋮	⋮	⋮		⋮		⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kh}	$T_{k.}$
Total	$T_{.1}$	$T_{.2}$	...	$T_{.j} = T_{.j}' + x$	...	$T_{.h}$	$G = G' + x$

where $T_{i.}'$ is total of known observations getting ith treatment.

$T_{.j}'$ is total of known observations in the jth block

G' is the total of all known observations.

$$\text{TSS} = \sum_{i=1}^{k}\sum_{j=1}^{h} y_{ij}^2 - \text{C.F.}$$

$$= x^2 + \text{constant w.r.t. } x - \text{C.F.}$$

$$\text{SST} = \frac{1}{h}\sum_{i=1}^{k} T_{i.}^2 - \text{C.F.}$$

$$= \frac{1}{h}(T_{i.}' + x)^2 + \text{constant w.r.t. } x - \text{C.F.}$$

$$\text{SSB} = \frac{1}{k}\sum_{j=1}^{h} T_{.j}^2 - \text{C.F.}$$

$$= \frac{1}{k}(T_{.j}' + x)^2 + \text{constant w.r.t. } x - \text{C.F.}$$

Now $$\text{C.F.} = \frac{G^2}{N} = \frac{(G' + x)^2}{N}$$

$\therefore$ $$\text{SSE} = E = \text{TSS} - \text{SST} - \text{SSB}$$

$$E = x^2 - \frac{1}{h}(T_{i.}' + x)^2 - \frac{1}{k}(T_{.j}' + x)^2 + \frac{(G' + x)^2}{N} + \text{constant w.r.t. } x$$

We choose the value of x such that E is minimum. To minimize E for variations in x by using principle of maxima and minima, we get

$$\frac{\partial E}{\partial x} = 0 \Rightarrow 2x - \frac{2(T_{i.}' + x)}{h} - \frac{2(T_{.j}' + x)}{k} + \frac{2(G' + x)}{hk} = 0$$

$$\Rightarrow \quad \left(1-\frac{1}{h}-\frac{1}{k}+\frac{1}{hk}\right)x-\frac{T_{i.}'}{h}-\frac{T_{j}'}{k}+\frac{G'}{hk}=0$$

$$\frac{hk-h-k-1}{hk}x-\left(\frac{k\,T_{i.}'+h\,T_{j}'-G'}{hk}\right)=0$$

$$(h-1)(k-1)\,x=k\,T_{i.}'+h\,T_{j}'-G'$$

$$x=\frac{k\,T_{i.}'+h\,T_{j}'-G'}{(h-1)(k-1)}$$

9.8. MISSING PLOT TECHNIQUE IN LSD

If any observation is missed due to some unknown reasons in LSD, we merely carry out the statistical analysis. The missing value is estimated by minimising the error sum of squares. This technique is called missing plot technique in LSD.

Let $y_{ijk} = x$ is missing in ith row, jth column by receiving the kth treatment in an $m \times m$ LSD.

Let R = Total of known observations in the ith row.

C = Total of known observations in the jth column.

T = Total of known observations due to the kth treatment.

S = Total of all known observations

Then

$$\text{TSS} = x^2 + \text{constant w.r.t. } x - \text{CF}$$

where $$\text{CF} = \frac{(\text{S} + x)^2}{m^2}$$

$$\text{SSR} = \frac{(\text{R} + x)^2}{m^2} + \text{constant w.r.t. } x - \text{CF}$$

$$SSC = \frac{(C + x)^2}{m^2} + \text{constant w.r.t. } x - CF$$

$$SST = \frac{(T + x)^2}{m^2} + \text{constant w.r.t. } x - CF$$

$$SSE = E = TSS - SSR - SSC - SST$$

$$E = x^2 - \frac{(R + x)^2}{m^2} - \frac{(C + x)^2}{m^2} - \frac{(T + x)^2}{m^2} + 2\,CF + \text{constant w.r.t. } x$$

$$= x^2 - \frac{1}{m}[(R + x)^2 + (C + x)^2 + (T + x)^2] + 2\,CF$$

We choose the value of x such that E is minimum. To minimise E for variations in x by using principle of maxima and minima, we get

$$\frac{\partial E}{\partial x} = 0$$

$$\Rightarrow \quad 2x - \frac{2}{m}[R + C + T + 3x] + 4\frac{S + x}{m^2} = 0$$

$$\Rightarrow \quad \frac{m^2x - m\,(R + C + T + 3x) + 2(S + x)}{m^2} = 0$$

$$\Rightarrow \quad (m^2 - 3m + 2)\,x - m\,(R + C + T) + 2S = 0$$

$$\Rightarrow \quad x = \frac{m\,(R + C + T) - 2S}{m^2 - 3m + 2}$$

$$x = \frac{m\,(R + C + T) - 2S}{(m - 1)\,(m - 2)}$$

8.7. EFFICIENCY OF RBD RELATIVE TO CRD

For k treatments and h replicates, we have the following table for RBD :

Source of variation	*d.f.*	*MSS*	*E (MSS)*
Treatments	$k-1$	S_T^2	$h\,\sigma_t^2 + \sigma_e^2$
Blocks	$h-1$	S_B^2	$k\,\sigma_b^2 + \sigma_e^2$
Error	$(k-1)(h-1)$	S_E^2	σ_e^2
Total	$kh-1$		

For the same experiment without blocks, we have the following table for CRD :

Source of variation	*d.f.*	*MSS*	*E (MSS)*
Treatments	$k-1$	$S_T'^2$	$h\,\sigma_t^2 + \sigma_e'^2$
Error	$N-k = hk-k$ $= k(h-1)$	$S_E'^2$	$\sigma_e'^2$
Total	$N-1 = hk-1$		

9.9. EFFICIENCY OF LSD

1. Relative Efficiency of LSD over RBD

For 'm' treatments, 'm' rows and 'm' columns, we have the following ANOVA table for LSD

Source of variation	*d.f.*	*MSS*	*E (MSS)*
Rows	$m-1$	S_R^2	$m\,\sigma_r^2 + \sigma_e^2$
Columns	$m-1$	S_C^2	$m\sigma_c^2 + \sigma_e^2$
Treatments	$m-1$	S_T^2	$m\,\sigma_t^2 + \sigma_e^{23}$
Error	$(m-1)(m-2)$	S_E^2	σ_e^2
Total	m^2-1		

(i) If we consider the same experiment is conducted with rows as blocks in RBD, we have the following ANOVA table for RBD.

Source of variation	*d.f.*	*MSS*	*E (MSS)*
Treatments	$m-1$	S_T^2	$m\,\sigma_t^2 + \sigma_e'^2$
Blocks	$m-1$	$S_B^2 = S_R^2$	$m\sigma_r^2 + \sigma_e'^2$
Error	$(m-1)\,(m-1) = (m-1)^2$	S_E^2	$\sigma_e'^2$
Total	m^2-1		

Since the total sum of squares remains same in each case, we get from the above tables

9.10. CONCEPT OF FACTORIAL EXPERIMENT

In the designs completely randomised design, randomised block design and Latin square design, only one factor is studied and analysed, *i.e.*, single set of treatments are compared and studied. Several factors of variation are studied and investigated simultaneously in factorial experiments. The study also concentrates on estimating the effects of each of the factors and analyse the variations in the effect of one factor to different levels of other factors.