

# **III B.Sc. Statistics**

**Subject name: Design of Experiment**

**Subject code :CST62**

**Unit                   :3**

## Completely Randomized Design [CRD]

In this design treatments are allocated at random to the experimental units over the entire experimental material.

Let us suppose that we have  $v$  treatments the  $i^{\text{th}}$  treatments being replication  $r_i$  times.

$$[i = 1, 2, \dots, v]$$

Then the whole experimental material is divided into  $n = \sum r_i$  experimental units and the treatments are distributed completely at random over the units subject to the condition that the  $i^{\text{th}}$  treatments occurs  $r_i$  times.

Randomisation assures that extraneous factors do not continually influence one treatment

$$r_i = r \quad \forall i = 1, 2, \dots, v$$

If each treatment is repeated and equal number of times & then  $n = r$  and randomisation gives every group of  $r$  units and equal chance of receiving

the treatments.

29/2

## Mathematical Model for CRD

The linear mathematical model:

$$x_{ij} = \mu_i + \varepsilon_{ij}$$

$$= \mu + (\mu_i - \mu) + \varepsilon_{ij}$$

$$= \mu + \alpha_i + \varepsilon_{ij}$$

where,

$\alpha_{ij} \rightarrow$  effect of  $i^{th}$  row and  $j^{th}$  column

$$(i = 1, \dots, k), (j = 1, 2, \dots, n_i)$$

$\mu \rightarrow$  General mean effect

$\alpha_i \rightarrow$  Effect due to  $i^{th}$  row (treatment)

$\varepsilon_{ij} \rightarrow$  Error effect due to chance

Null hypothesis:  $H_0$

To test the equality of the population means

(i.e.) all mean values are homogeneity

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k = \mu \text{ which dedu}$$

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

Alternative hypothesis ( $H_1$ )

All mean values are not homogeneity  
(Heterogeneity)

$$H_1: \mu_1 \neq \mu_2 \dots \neq \mu_k \neq \mu$$

Statistical analysis:

We consider

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.} + \bar{x}_{i.} - \bar{x}_{..})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} [(x_{ij} - \bar{x}_{i.})^2 + (\bar{x}_{i.} - \bar{x}_{..})^2 \\ &\quad + 2(x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..})] \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x}_{..})^2 \\ &\quad + 2 \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) \end{aligned}$$

Since the product term is vanished

$$2 \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) = 0$$

The algebraic sum of deviation of the  $i^{th}$  class from their mean is zero.

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 + \\ &\quad \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x}_{..})^2 \end{aligned}$$

Alternative hypothesis, H<sub>1</sub>

All mean values are not homogeneity  
(Heterogeneity)

$$H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_k \neq \mu$$

Statistical analysis:

Take constant

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{ij} + \bar{x}_{ij} - \bar{x}_{..})^2$$

$$= \sum_{j=1}^k \sum_{i=1}^{n_j} [(x_{ij} - \bar{x}_{ij})^2 + (\bar{x}_{ij} - \bar{x}_{..})^2 + 2(x_{ij} - \bar{x}_{ij})(\bar{x}_{ij} - \bar{x}_{..})]$$

$$= \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{ij})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x}_{ij} - \bar{x}_{..})^2 + 2 \sum_{j=1}^k \sum_{i=1}^{n_j} [(\bar{x}_{ij} - \bar{x}_{ij})(\bar{x}_{ij} - \bar{x}_{..})]$$

Since the product term is vanished

$$2 \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x}_{ij} - \bar{x}_{ij})(\bar{x}_{ij} - \bar{x}_{..}) = 0$$

The algebraic sum of deviation of the  $i^{th}$  class from their mean is zero.

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{ij})^2 +$$

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x}_{ij} - \bar{x}_{..})^2$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 = \text{Total sum of squares}$$

$$= TSS \text{ or } ST^2$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{ij})^2 = \text{error sum of squares}$$

$$= SSE \text{ or } SE^2$$

$$\sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{..})^2 = \text{column sum of squares}$$

or sum of square due to

treatment

$$= SSC \text{ or } SC^2$$

$$TSS / \text{Total sum of squares} = SSC + SSE$$

$$ST^2 = SC^2 + SE^2$$

Degrees of freedom:

Degrees of freedom for column c-1

Degrees of freedom for error N-c

Degrees of freedom for total N-1

Mean sum of squares:

Mean sum of squares for treatment  
or column

$$MSSc = \frac{SSC}{c-1} = \frac{SC^2}{c-1}$$

Mean sum of square for error

$$MSSE = \frac{SSE}{N-C} = \frac{SE^2}{N-C}$$

### ANOVA - Table

source of variation	sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Between column	$\sum n_i (\bar{x}_{ij} - \bar{x}_{..})^2$ $SSC (n\bar{x})$ $SC^2$	$N-C = 1$	$MSSc = \frac{SC^2}{N-C}$ $or$ $SC^2$	
Within	$\sum (x_{ij} - \bar{x}_{..})^2$ $SSE (n\bar{x})$ $SE^2$	$N-C = N-1$	$MSSE = \frac{SE^2}{N-C}$	$F_c = \frac{SC^2}{SE^2}$
Total	$\sum \sum (x_{ij} - \bar{x}_{..})^2$ $= ST^2$	$N-1$		

If calculated value is greater than the table value at 1% or 5% level

of significance the  $H_0$  is rejected  
otherwise we accept  $H_0$ .

### Advantages of CRD

\* CRD results in the maximum use of the experimental units, since all the experimental material can be used.

\* The design is very flexible. Any number of treatments can be used. Different treatments can be used unequal number of times without unduly complicating the statistical analysis in most of the cases.

\* It provides the maximum number of degrees of freedom for the estimating of the error variance which influences the sensitivity or the precision of the experiments for small experiments.

\* The statistical analysis remains simple if some or all the observations for any treatment are rejected or lost or missing for some purely random accidental reasons.

### Disadvantages of CRD

\* The design suffers from inherently less informative than other more sophisticated layouts. This usually happens if the experimental material is not homogeneous.

\* The design loss efficiency and results in loss sensitivity in detecting significant effect.

\* Due to the fertility gradient of the soil the whole experimental material in field is not homogeneous and it is better to use more efficient designs like RBD or LSD.

### Applications

\* CRD is most useful in laboratory technique and methodological studies

Example:

Physics, chemistry, cookery  
in chemical and biological experiments.

\* CRD is also recommended in situation where an appreciable fraction of units is likely to be destroyed or to fail respond.

### Randomised Block Design [RBD]

In field experimentation if the whole of the experimental area is not homogeneous and fertility gradient is only in one direction, then a simple method of controlling the variability

of the experimental material ~~consid~~  
in stratifying or grouping the ~~cons~~  
area into relatively homogeneous  
strata or subgroups, perpendicular  
to the direction of the fertility  
gradient.

If the treatments are applied  
random to relatively homogeneous  
units within each strata or block  
and replicated overall the blocks  
is known as Randomised Block Design.

### 1/3 Layout of RBD

Let us consider five treatments  
A, B, C, D, E each replicated four times.  
the layout will be

Block I	A	E	B	D	C
Block II	E	D	C	B	A
Block III	C	B	A	E	D
Block IV	A	D	E	C	B

Random permutation

In Block we allocate the treat-

A - 1<sup>st</sup> plot

B - 3<sup>rd</sup> plot

C - 5<sup>th</sup> plot

D - 4<sup>th</sup> plot

E - 2<sup>nd</sup> plot

Similarly, we allocate the treatments for all the blocks.

Mathematical model for RBD

Let us suppose that  $\alpha_{ij}$  ( $i=1, 2, \dots, k$ ,  $j=1, 2, \dots, h$ ) are independent the linear mathematical model

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where,

$x_{ij} \rightarrow$  Effect of  $i^{\text{th}}$  row and  $j^{\text{th}}$  column

$\mu \rightarrow$  General mean effect

$\alpha_i \rightarrow$  effect due to  $i^{\text{th}}$  row

$\beta_j \rightarrow$  effect due to  $j^{\text{th}}$  column

$\varepsilon_{ij} \rightarrow$  Error effect due to chance

Null hypothesis ( $H_0$ )

The treatment as well as varieties

are homogeneous

$$H_0(t): \mu_{1\cdot} = \mu_{2\cdot} = \dots = \mu_{k\cdot} = \mu$$

$$H_0(v): \mu_{\cdot 1} = \mu_{\cdot 2} = \dots = \mu_{\cdot h} = \mu$$

these equivalent

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

$$H_1 : \beta_1 = \beta_2 = \dots = \beta_h = 0$$

Statistical analysis:

We can consider

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^h (\bar{x}_{ij} - \bar{x}..)^2 &= \sum \sum [\bar{x}_{ij} - \bar{x}_{i.} + \bar{x}_{i..} \\ &\quad + \bar{x}_{.j} - \bar{x}.. + \bar{x}.. - \bar{x}]^2 \\ &= \sum \sum [\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..] + \\ &\quad (\bar{x}_{i.} - \bar{x}..) + (\bar{x}_{.j} - \bar{x}..)]^2 \\ &= \sum \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..)^2 + \\ &\quad \sum \sum (\bar{x}_{i.} - \bar{x}..)^2 + \sum \sum (\bar{x}_{.j} - \bar{x}..)^2 \\ &\quad + 2 \sum \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..) (\bar{x}_{i.} - \bar{x}..) \\ &\quad + 2 \sum \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..) (\bar{x}_{.j} - \bar{x}..) \\ &\quad + 2 \sum \sum (\bar{x}_{i.} - \bar{x}..) (\bar{x}_{.j} - \bar{x}..) \end{aligned}$$

Since algebraic sum of deviation set of observation about their mean is zero that is all the product term is vanished. So, we have

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^h (\bar{x}_{ij} - \bar{x}..)^2 &= \sum \sum (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}..)^2 \\ &\quad + h \sum_{i=1}^k (\bar{x}_{i.} - \bar{x}..)^2 \\ &\quad + k \sum_{j=1}^h (\bar{x}_{.j} - \bar{x}..)^2 \end{aligned}$$

$$TSS = SSR + SSc + SSE$$

$$ST^2 = SR^2 + SSc^2 + SE^2$$

where

- \*  $\sum_{i=1}^k \sum_{j=1}^h (x_{ij} - \bar{x}_{..})^2$  is the total sum of squares  
= TSS or  $ST^2$

- \*  $\sum_{i=1}^h (\bar{x}_{i..} - \bar{x}_{..})^2$  is row sum of squares  
= SSR or  $SR^2$
- \*  $\sum_{j=1}^h (\bar{x}_{.j} - \bar{x}_{..})^2$  is column sum of squares  
= SSc or  $Sc^2$

- \*  $\sum_{i=1}^k \sum_{j=1}^h (x_{ij} - \bar{x}_{i..} - \bar{x}_{.j} + \bar{x}_{..})^2$  is error sum of squares  
= SSE or  $SE^2$

Degrees of freedom:

For row  $\gamma = r - 1$

For column =  $c - 1$

For error  $\gamma = (r-1)(c-1)$

For total  $\gamma = (N-1)$  or  $rc - 1$

## Mean sum of square

The sum of square divided by its degrees of freedom is known as mean sum of square.

\* Mean sum of square for row:

$$SR^2 = \frac{SSR}{r-1} \quad (\text{or}) \quad \frac{SR^2}{r-1}$$

\* Mean sum of square for column:

$$SC^2 = \frac{SSC}{c-1} \quad \text{or} \quad \frac{SC^2}{c-1}$$

\* Mean sum of square for error:

$$SE^2 = \frac{SSE}{(r-1)(c-1)} \quad \text{or} \quad \frac{SE^2}{(r-1)(c-1)}$$

Test statistic

For row

$$FR = \frac{SR^2}{SE^2}$$

For column

$$Fc = \frac{SC^2}{SE^2}$$

# ANOVA table:

SOURCE of variation	sum of squares	Degrees of freedom	mean sum of squares	Variate ratio
between rows	$SBR = h \sum (\bar{x}_{1j} - \bar{x}_{..})^2$	$(r-1)$	$BR^2 = \frac{SBR^2}{r-1}$	$F_R = \frac{MSSR}{MSSC}$
Between columns	$SSC = K \sum (\bar{x}_{.j} - \bar{x}_{..})^2 (C-1)$	$C-1$	$SC^2 = \frac{SSC^2}{C-1}$	$F_C = \frac{MSSC}{MSSE}$
Total	$TSS = \sum (x_{ij} - \bar{x}_{..})^2$	$N-1$		

If an observed value of  $F >$  than the tabulated value of  $F_\alpha$  at specified level of significance that is 5% or 1%, the  $H_0$  is rejected.

Advantages: [RBD]

i) Accuracy:

This design has been shown to be more efficient or accurate than completely Randomised Design for the most types of experimental work.

### 2) Flexibility:

In RBD no restrictions are placed on the number of treatments or the number of replicates. In general at least two replicates are required to carry out the test of significance.

### 3) Ease of analysis:

Statistical analysis is simple and rapid. The error of any treatment can be isolated and any number of treatments may be omitted from the analysis without complicating it.

### Disadvantages of RBD

Randomised Block Design is not suitable for large number of treatments as for cases in which complete block contains considerable variability.

### Difference between CRD and RBD

In CRD we do not resort to the grouping of the experimental site (space, material, time) and allocate the treatments at random to the

experimental units.

But In RBD treatments are allocated at random within the units of each stratum or block & randomization is maintained. Also variation among blocks is removed from variation due to error.

### Characteristics of RBD

- \* No interaction between treatments and blocks. Interaction will be separable from experimental error and if the interaction are large, the experiment may yield misleading results.
- \* It is flexible readily adaptable and easy to analyse and these points have made it the most popular of all the designs with LSD.
- \* RBD provides a method of eliminating or reducing the effects of trends.
- \* All the treatments are to be applied within each block, In each block we take many units as the number of treatments. Each treatment will have the same number of replications.

11/3

## Latin square Design [LSD]

A useful method of eliminating fertility variations consist in an experimental layout which will control variation in two perpendicular directions. Such a layout is a Latin Square Design.

Layout of Design: start  $4 \times 4 - 12 \times 12$

In this design a number of treatments is equal to the number of replicates.

Eg: In case of  $m$  treatments,  $m$  replicates there have to be  $m \times m = m^2$  experimental units. The whole of experimental area is divided into  $m^2$  experimental units arranged in a square, so that each row as well as each column contains  $m$  units.

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

# Mathematical model for LSD

## The linear model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}$$

where

$Y_{ijk} \rightarrow$  is the  $i^{th}$  row  $j^{th}$  column

and receiving the  $k^{th}$  treatment.

$\mu \rightarrow$  general mean effect

$\alpha_i \rightarrow$  effect due to  $i^{th}$  row

$\beta_j \rightarrow$  effect due to  $j^{th}$  column

$\tau_k \rightarrow$  effect due to  $k^{th}$  treatment

$\varepsilon_{ijk} \rightarrow$  <sup>error</sup> effect due to random

component assumed to be normally distributed.

Null hypothesis:  $H_0$

$$H_\alpha : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$H_\beta : \beta_1 = \beta_2 = \dots = \beta_m = 0$$

$$H_T : \tau_1 = \tau_2 = \dots = \tau_m = 0$$

Statistical analysis of  $t$  &  $SD$ :

We consider:

$$\begin{aligned}\sum_{i,j,k \in S} (y_{ijk} - \bar{y}_{...})^2 &= \sum_{i,j,k \in S} (y_{ijk} + \bar{y}_{...} - \bar{y}_{...} - \bar{y}_{...})^2 \\ \bar{y}_{...} &= \bar{y}_{...} + \bar{y}_{...k} - \bar{y}_{...k} + 2\bar{y}_{...} - 2\bar{y}_{...} \\ &= \sum_{i,j,k \in S} [(\bar{y}_{...} - \bar{y}_{...}) + (\bar{y}_{...j} - \bar{y}_{...}) + (\bar{y}_{...k} - \bar{y}_{...}) \\ &\quad + (y_{ijk} - \bar{y}_{...} - \bar{y}_{...j} - \bar{y}_{...k} + 2\bar{y}_{...})]^2 \\ &= m \sum_i (\bar{y}_{...} - \bar{y}_{...})^2 + m \sum_j (\bar{y}_{...j} - \bar{y}_{...})^2 + \\ &\quad m \sum_k (\bar{y}_{...k} - \bar{y}_{...})^2 + \sum_{i,j,k \in S} (y_{ijk} - \bar{y}_{...} - \bar{y}_{...j} - \\ &\quad - \bar{y}_{...k} + 2\bar{y}_{...})^2\end{aligned}$$

The product term vanish since the algebraic sum of deviations from mean will zero.

Where

- \*  $\sum_{i,j,k \in S} (y_{ijk} - \bar{y}_{...})^2$  = is total sum of square
- \*  $m \sum_i (\bar{y}_{...} - \bar{y}_{...})^2$  = Row sum of square
- \*  $m \sum_j (\bar{y}_{...j} - \bar{y}_{...})^2$  = Column sum of square

$$m \sum_k (\bar{y}_{..k} - \bar{y}...)^2 = \text{varieties sum of square}$$

$$\sum_{i,j,k \in S} (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}...)^2 = \text{Error sum of square}$$

H3

Degrees of freedom:

Degrees of freedom for row  $\tau = m - 1$

Degrees of freedom for column  $\tau = m - 1$

Degrees of freedom for treatment  $\tau = m - 1$

Degrees of freedom for error  $\tau = (m-1)(m-2)$

Mean sum of square:

$$\text{For row, } DR^2 = \frac{SR^2}{m-1} \text{ (or) } \frac{SSR}{m-1}$$

$$\text{For column, } SC^2 = \frac{SC^2}{m-1} \text{ (or) } \frac{SSC}{m-1}$$

$$\text{For treatment, } SV^2 = \frac{SV^2}{m-1} \text{ (or) } \frac{SSV}{m-1}$$

$$\text{For error, } DE^2 = \frac{SE^2}{m-1} \text{ (or) } \frac{SSE}{(m-1)(m-2)}$$

Test statistic:

$$\text{For row, } F_R = \frac{SR^2}{SE^2}$$

$$\text{For column, } F_C = \frac{SC^2}{SE^2}$$

$$\text{For variety, } F_V = \frac{SV^2}{SE^2}$$

### ANOVA - Table

Source of Variation	Sum of Squares	Degrees of freedom	Mean Sum of Square	Variation
For row	$m \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$m-1$	$SR^2 = \frac{SR^2}{m-1}$	$F_R = \frac{SR^2}{SE^2}$
For column	$m \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	$m-1$	$SC^2 = \frac{SC^2}{m-1}$	$F_C = \frac{SC^2}{SE^2}$
For varieties	$m \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$m-1$	$SV^2 = \frac{SV^2}{m-1}$	$F_V = \frac{SV^2}{SE^2}$
For error	$\sum_{i,j,k} (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$	$(m-1)(m-2)$	$SE^2 = \frac{SE^2}{(m-1)(m-2)}$	$F_V = \frac{SE^2}{SE^2}$
Total	$\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2$	$N-1$		

Advantages of LSD:

\* LSD controls more of the variation than CRD and RBD.

\* LSD is an incomplete three way layout then its advantage over the complete three way layout is that

instead of  $m^3$  experimental units  
 $m^2$  units are needed.

\* Thus a  $4 \times 4$  LSD results in  
saving of  $64 - 16 = 48$  observations over a  
complete three-way layout.

\* Statistical analysis is simple, though  
slightly complicated than for RBD. Even  
with missing data the analysis remains  
relatively simple.

\* More than one factor can be  
investigated simultaneously and with  
less trials than more complicated design.

### Disadvantages of LSD:

The fundamental assumption that  
there is no interaction between  
different factors may not be true in  
general.

Unlike RBD in LSD the no. of  
treatment is restricted to the replications  
and this limits its field of applications.

LSD is suitable for the no. of treatments  
between 5 and 10 and for more than  
10-12 treatments the design is seldom

used.

- \* In case of missing plots when several units are missing, the statistical analysis becomes quite complex.

In the field layout RBD is much easier to manage than LSD. Since the former can be performed equally well on a square or rectangular field or a field of any shape whereas for the later approximately a square field is necessary.