

30 pg



# DESIGN OF EXPERIMENT



## CHAPTER OUTLINE

After studying this chapter, students will be able to understand the:

- ❖ Experimental design
- ❖ Basic Principles of experimental designs
- ❖ Completely Randomized Design (CRD), Randomized Block Design (RBD), ANOVA table, Efficiency of RBD relative to CRD, Estimations of missing value (one observation only), Advantages and disadvantages
- ❖ Latin Square Design (LSD), Statistical analysis of  $m \times m$  LSD for one observation per experimental unit,
- ❖ ANOVA table, Estimation of missing value in LSD (One observation only)
- ❖ Efficiency of LSD relative to RBD, Advantage and disadvantages
- ❖ Problems and illustrative examples related using software.

## **Experimental Design**

In any experiment the prime objective is either to maximize the precision or to minimize the error. Hence the validity of the experiment depends upon how adequately it has been designed. It helps to determine the relationship between causes and effects of the various experimental treatments. The experiments are carried out under controlled conditions so that input and output can be measured directly.

### **Objectives of the Experimental Design**

- i) To estimate the effects of various treatments and to compare the differences of effects are significant or not.
- ii) To estimate the interaction effects of various treatments and to compare them
- iii) To establish the mathematical relationship between various treatments and their effects.
- iv) To estimate error
- v) To control error
- vi) Proper interpretation of the results

### **Steps in Design of Experiments**

The major steps in the experimental design are

- i) The problem which has to be solved by experiment should be stated in clear and significant way.
- ii) The hypothesis in accordance with the experiment should be formulated in a clear way.
- iii) The experimental technique and designing to solve the problem should be divided in a systematic way.
- iv) The result that come from the designing should be examined carefully and reference back to the reasons for the enquiry of the required information. This is done to ensure that experiment provides an adequate extent.
- v) The possible outcomes that come from the experiment should be under the consideration of statistical procedure. This is done to ensure the validity of the conditions necessary for statistical procedure.
- vi) The conclusion is drawn after the experiment with the measure of reliability of the estimate of any treatment that is evaluated.
- vii) The whole investigation is evaluated specially with the other investigation on the same or similar problems.

### **Terminology in Experimental Design**

#### **Experiment**

It is means of getting an answer to a question that the experimenter has in mind. In planning experiment, we clearly state our objectives and formulate the hypothesis we want to test. Experiment can be divided into two categories i) Absolute and ii). Comparative Absolute

experiment consists of determining the absolute value of some characteristic such as finding correlation coefficient between two variable, average intelligence of a group of people etc. While the comparative experiment consists of comparing different types of fertilizers, different types of cultivation methods, different varieties of crops etc.

### Treatment

These are the inputs whose outcomes are to be estimated and compared. These are the different procedures under comparison in the experiment. In agricultural experiment different types of fertilizers, different types of cultivation process, different varieties of crops are treatments.

### Experimental Unit

The smallest division of the experimental material in which the treatments are applied and the effects of treatments are measured. In agricultural experiment plot are the experimental unit. In feeding experiment of cows, a cow is experimental unit. Among the patients admitted in a hospital, a patient is an experimental unit.

### Yields (Effects)

The outcomes of the experiment due to the application of treatments in experimental units are called yields. In agricultural experiment production of crop on using different fertilizers are yields.

### Blocks

The experimental field is divided into relatively homogeneous subgroups or strata which is homogeneous or uniform among themselves than the field as a whole are called blocks.

### Experimental Error

A fundamental phenomenon in replicated experiments is the variation in the measurements made on different experimental units even when they get the same treatment. A part of this variation is systematic and can be explained whereas the remainder is to be taken of the random type. This unexplained random part of the variation is termed as experimental error. This is a technical term and does not mean a mistake but includes all types of extraneous variation due to (i) inherent variability in the experimental units, (ii) small errors associated with the measurements made and (iii) lack of representativeness of the sample to the population under study.

It provides a basis for the confidence to be placed in the inference about the population. So it is to be estimated and controlled. It can be estimated by the replication and can be controlled by use of local control.

### Precision

It is the amount of information or sensitivity of an experiment. It is given by reciprocal of the variance of the mean. If an experiment is replicated  $r$  times the precision of experiment is

$$\frac{1}{\text{Var}(\bar{x})} = \frac{r}{\sigma^2}, \sigma^2 \text{ is variance.}$$

The precision increases as the replication increases or the variance decreases.

**Efficiency of Design**

Consider any two design  $D_1$  and  $D_2$  with replications  $r_1$  and  $r_2$  and variance  $\sigma_1^2$  and  $\sigma_2^2$  respectively, then the ratio of the precisions of design  $D_1$  and  $D_2$  is called the relative efficiency of design  $D_1$  with respect to  $D_2$ . It is given by  $E = \frac{r_1}{s_1^2} / \frac{r_2}{s_2^2}$ .

If  $E = 1$ , both the design  $D_1$  and  $D_2$  are equally efficient.

If  $E > 1$ , design  $D_1$  is more efficient than design  $D_2$ .

If  $E < 1$ , design  $D_1$  is less efficient than design  $D_2$ .

**Basic principles of experimental design**

Designing of an experiment is deciding how the observations should be taken to answer a particular question in a valid, efficient and economic way. The design and analysis go together in the sense that if an experiment is properly designed then there will exist an appropriate way of analyzing the data. The application of the technique of analysis of variance is appropriate only when the data conform to the basic set up of the analysis of variance. The analysis of the data will be meaningless if the assumption in the analysis of variance are not fulfilled. Hence the layout and analysis of data are coordinated in the design of experiments.

According to R.A. Fisher a good experimental design must posses the following three principles namely;

- (i) Replication (ii) Randomization (iii) Local control

**Replication**

It is the repetition of treatments under investigation. A treatment is repeated a large number of times in order to obtain more reliable result than is possible from single observation.

It works in two ways:

- i) Along with randomization, it provides an estimate of the treatment effect.
- ii) Along with local control, it provides minimization of error.

The most effective way to increase precision is to increase the number of replication. The precision can be increased by increasing the plot size but it has found that increasing replication of small plot is more efficient than using larger plot. The replication in a particular case depends on the variability of the material, cost of taking observations etc. It broadens the scope of experiment by including different types of experimental units.

**Randomization**

It is a process of allocating treatments to various plots in a random manner. It ensures that each treatment will have an equal chance of being assigned to an experimental unit. It greatly reduces the bias of applying a particular treatment to a particular unit. It is essential for valid estimate of experimental error and to minimize bias in the results. It is precaution against disturbances (errors) that may or may not occur and may or may not be serious if occur.

- Its aims:
- Equalization of factor not under control
  - Eliminates bias in any form
  - Provides a basis to estimate the treatment effect.

### Local control

If the experimental material is heterogeneous and different treatments are allocated to various experimental units (plots) in random manner then experimental error will be increased. It is desirable to reduce the experimental error without increasing replication or without interfering the required randomness.

The experimental error can be minimized by making the relatively heterogeneous experimental material into relatively homogeneous blocks is called local control. For this purpose the whole experimental area(material) is divided into a number of blocks perpendicular to the direction of fertility gradient the objective of local control is to give equal advantage of soil fertility to all the treatments. The experimental material is divided into a number of block rowwise or columnwise or both such that variation between blocks is maximum and variation within each block is minimum. Divide each block into as many plots as the number of treatments and allocate the treatments to the plots of each block separately in a random manner.

### Completely randomized design (CRD)

It is simplest of all the design which is based upon only two principles of design namely replication and randomization. In this design treatments are assigned completely at random manner so that each and every experimental unit has equal chance of receiving any treatment. it is appropriate for the homogeneous experimental material.

#### Layout

Let us consider  $t$  treatments in which  $i^{\text{th}}$  treatment is replicated  $r_i$  times,  $i = 1, 2, 3, 4, \dots, t$  so that

there are  $n$  experimental units such that  $n = \sum_{i=1}^t r_i$ . The homogeneous experimental material is

divided into  $n$  plots(experimental units) and  $i^{\text{th}}$  treatment repeats in  $r_i$  units. The treatments are replicated along row wise as well as column wise. We study the variation between treatments. Observations are classified according to one way. Hence this is the case of one way ANOVA.

Consider a particular case in which  $r_1 = r_2 = r_3 = \dots = r_t = r$  then  $n = rt$ . In this case each treatment is repeated equal number of times  $r$ . In general equal number of replication should be made for each treatment except in the case when some treatments are of greater interest than the others. For example, consider  $t = 3$  (A, B, C) and  $r = 4$  then the treatments are allocated as shown below;

C	A	B	A
C	B	A	A
B	B	C	C

**Mathematical Model**

$$y_{ij} = \mu + \tau_i + e_{ij}; i = 1, 2, 3, \dots, t; j = 1, 2, 3, \dots, r$$

Where,

$y_{ij}$  =  $j^{\text{th}}$  unit receiving  $i^{\text{th}}$  treatment

$\mu$  = general mean effect

$\tau_i$  = effect due to  $i^{\text{th}}$  treatment

$e_{ij}$  = error due to chance

**Assumptions**

- i) All the observations are independent.
- ii) All the observations are drawn from population having constant variance.
- iii) All the treatments should be homogeneous as far as possible.
- iv) Various treatments and environmental effects should be additive in nature.
- v) All  $e_{ij}$  are i.i.d.  $N(0, \sigma_e^2)$

**Problem to test**

$H_0$ :  $\mu_1 = \mu_2 = \mu_3 = \mu_t$ . (There is no significant difference between the treatments)

$H_1$ : At least one  $\mu_i$  is different.  $i = 1, 2, 3, \dots, t$  (There is at least one significant difference between treatments)

**Statistical Analysis**

In the model  $y_{ij} = \mu + \tau_i + e_{ij}$ , parameters  $\mu$  and  $\tau_i$  are determined by using the principle of least square by minimizing error(residual) sum of square.

Total sum of square (TSS) = Sum of square due to treatment (SST) + Sum of square due to error (SSE)

**Degree of freedom (d.f.) for various sum of square**

Degree of freedom for total sum of square =  $rt - 1 = n - 1$

Degree of freedom for sum of square due to treatment =  $t - 1$

Degree of freedom for sum of square due to error =  $n - t$

**Mean Sum of Square (MSS)**

The sum of square divided by the corresponding degree of freedom gives the respective mean sum of square or variance.

Mean sum of square due to treatment (MST) =  $SST / (t - 1)$

Mean sum of square due to error (MSE) =  $SSE / (n - t)$

**ANOVA table**

S.V.	d.f.	S.S.	M.S.	$F_{\text{cal}}$	$F_{\text{tab}}$
Treatment	$t - 1$	SST	$MST = SST / (t - 1)$	$F_T = MST / MSE$	$F_{\alpha / (t-1), (n-t)}$
Error	$t(r - 1)$	SSE	$MSE = SSE / (n - t)$		
Total	$r(t - 1)$	TSS			

**Decision**

Reject  $H_0$  at  $\alpha\%$  level of significance if  $F_T > F_{\alpha}((t-1), (r-1))$ , accept otherwise.

**Relation to calculate TSS, SST and SSE**

$$TSS = \sum_{i=1}^t \sum_{j=1}^r (\bar{y}_{ij} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - C.F.$$

$$SST = r \sum_{i=1}^t (y_i - \bar{y}_{..})^2$$

$$= \frac{\sum_{i=1}^t T_i^2}{r} - C.F. \text{ where } C.F. = \frac{G^2}{N}$$

$$SSE = TSS - SST$$

**Advantages of CRD**

- It is easy to layout,
- It allows maximum number of d.f. for MSS due to error which minimize error sum of square.
- It is simple to statistical analysis due to one way classification.
- If some observations are missing the analysis still remains simple.

**Disadvantages of CRD**

- Principle of local control is not used.
- It is suitable for small treatments, replication and homogeneous material only.

**Uses of CRD**

- It is used in green house, laboratory etc.

**Example 1:** The yield of treatments in different plots are as shown below. Carry out analysis.

t <sub>4</sub> 1401	t <sub>3</sub> 2536	t <sub>3</sub> 2459	t <sub>1</sub> 2537	t <sub>3</sub> 2827	t <sub>1</sub> 2069
t <sub>2</sub> 2211	t <sub>1</sub> 1797	t <sub>4</sub> 1170	t <sub>4</sub> 1516	t <sub>4</sub> 2104	t <sub>3</sub> 2385
t <sub>2</sub> 3366	t <sub>1</sub> 2104	t <sub>2</sub> 2591	t <sub>3</sub> 2460	t <sub>4</sub> 1077	t <sub>2</sub> 2544

**Solution:**

Problem to test

 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ . $H_{11}: \text{At least one of the } \mu_i \text{ is different, } i = 1, 2, 3, 4 \quad (1 = t_1, 2 = t_2, 3 = t_3, 4 = t_4)$

						$T_i$
$t_1$	2537	2069	1797	2104		6049
$t_2$	2211	3366	2591	2544		10712
$t_3$	2536	2459	2827	2385	2460	12667
$t_4$	1401	1170	1516	2104	1077	6427
						$\Sigma T_i = 35855$

Now,  $\sum_{i,j} y_{ij}^2 = 2537^2 + 2069^2 + 1797^2 + 2104^2 + \dots + 1077^2 = 88374337$

$$G = \Sigma T_{..} = 35855$$

$$N = n_1 + n_2 + n_3 + n_4 = 4 + 4 + 5 + 5 = 18$$

$$C.F. = \frac{G^2}{N} = \frac{(35855)^2}{18} = 71421168.06$$

$$\begin{aligned} TSS &= \sum_{i,j} y_{ij}^2 - C.F. \\ &= 88374337 - 71421168.06 \\ &= 16953168.94 \end{aligned}$$

$$\begin{aligned} SST &= \sum_i \frac{T_i^2}{n_i} - C.F. \\ &= \left\{ \frac{(6049)^2}{4} + \frac{(10712)^2}{4} + \frac{(12667)^2}{5} + \frac{(6427)^2}{5} \right\} - 71421168.06 = 5768691.79 \end{aligned}$$

$$\begin{aligned} SSE &= TSS - SST \\ &= 16953168.94 - 5768691.79 = 11184477.15 \end{aligned}$$

### ANOVA table

S.V.	d.f.	S.S.	M.S.	$F_{Cal}$	$F_{Tab}$
Treatment	3	5768691.79	1922897.26	22.72	$F_{0.05(3,14)} = 5.66$
Error	14	11184477.15	84605.51		
Total	17	16953168.94			

### Decision

$F_T = 22.72 > F_{0.05(3,14)} = 5.66$ , reject  $H_0T$  at 5% level of significance.

### Conclusion

Treatment difference is highly significant.

*Example 2:*

There are different types of text entry techniques A, B and C. The efficiency is measured by its error rate. The error rate found while entering in four different mobile are given below.

A	8	C	5	C	4	B	5
B	7	A	4	B	1	A	6
C	2	B	5	C	3	A	7

Test the significant difference among the three text entry techniques.

**Solution:**

Problem to test

H<sub>0</sub>: The difference in efficiency among the text entry technique are not significant.

H<sub>1</sub>: The difference in efficiency among the text entry technique are significant.

Now,

A	8	4	6	7
B	7	5	5	1
C	2	5	4	3

Changing origin to 4 ( $u_{ij} = x_{ij} - 4$ )

Text entry technique					T <sub>i..</sub>
A	4	0	2	3	9
B	3	1	1	-3	2
C	-2	1	0	-1	-2
					$\Sigma T_{i..} = 9$

Here,

$$G = \Sigma T_{i..} = 9, N = 12$$

$$C.F. = \frac{G^2}{N} = \frac{(9)^2}{12} = 6.75$$

$$\sum_{i,j} u_{ij}^2 = 16 + 0 + 4 + 9 + 9 + 1 + 1 + 9 + 4 + 1 + 0 + 1 = 55$$

$$TSS = \sum_{i,j} u_{ij}^2 - C.F. = 55 - 6.75 = 48.25$$

$$SST = \frac{\sum_i T_{i..}^2}{r} - C.F.$$

$$= \frac{1}{4} (81 + 4 + 4) - 6.75 = 22.25 - 6.75 = 15.5$$

$$SSE = TSS - SST$$

$$= 48.25 - 15.5 = 32.75$$

ANOVA table

S.V.	d.f.	S.S.	M.S.	F <sub>Cal</sub>	F <sub>Tab</sub>
Treatment	2	15.5	7.75	2.73	F <sub>0.05(2,9)</sub> = 4.26
Error	9	32.75	3.63		
Total	11	48.25			

**Decision**

$F_T = 2.73 < F_{0.05(2,9)} = 4.26$ , accept  $H_0$  at 5% level of significance.

**Conclusion**

The efficiency of three techniques are not significant.

**Randomised Block Design (RBD)**

When the experimental material is not homogeneous the RBD is better than CRD. The RBD is the design where the treatments are allocated in random manner but randomization is restricted that each treatment must occur once in each row or once in each column. Hence this design is row wise or column wise. It is based upon the all principles of design namely replication, randomization and local control.

**Lay out**

Let us consider  $t$  treatments with  $r$  replication each so that there are  $N = r t$  experimental unit.

Let us divide the non homogeneous experimental material into a number of relatively homogeneous blocks perpendicular to the direction of fertility gradient. Divide each block into as many plot as the number of treatment and allocate these  $t$  treatments randomly to the plot of each block separately.

**The treatments are replicated along row wise or column wise.** In this case we study the variation between treatments and the variation between blocks. Observation are classified according to two ways. Hence it is the case of two way ANOVA.

Let us consider  $t = 3(A, B, C)$  and  $r = 4$ . The treatments are allocated in the blocks as shown below.

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

**Mathematical model**

$$Y_{ij} = \mu + \tau_i + \beta_j + e_{ij}$$

Where,

$y_{ij}$  =  $j^{\text{th}}$  block receiving  $i^{\text{th}}$  treatment.

$i = 1, 2, \dots, t, j = 1, 2, \dots, r$

$\mu$  = constant effect

$\tau_i$  = Effect due to  $i^{\text{th}}$  treatment

$\beta_j$  = Effect due to  $j^{\text{th}}$  block

$e_{ij}$  = Error due to chance

**Assumptions**

- (i) All the observations are independent.
- (ii) All the observations should be drawn from normal population having constant variance.
- (iii) All the treatments should be homogeneous as far as possible.
- (iv) Various treatments and environmental effect are additive in nature.
- (v)  $e_{ij}$  are i.i.d.  $N(0, \sigma_e^2)$

**Statistical analysis**

In the model  $y_{ij} = \mu + \tau_i + \beta_j + e_{ij}$ , parameters  $\mu$ ,  $\tau_i$  and  $\beta_j$  are determined by using the principle of least square by minimizing error(residual) sum of square.

Total sum of square (TSS) = Sum of square due to treatment (SST) + Sum of square due to block (SSB) + Sum of square due to error (SSE)

**Degree of freedom (d.f.) for various sum of square:**

Degree of freedom for total sum of square =  $rt - 1 = N - 1$

Degree of freedom for sum of square due to treatment =  $t - 1$

Degree of freedom for sum of square due to block =  $r - 1$

Degree of freedom for sum of square due to error =  $t(r - 1)$

**Mean Sum of Square (MSS):**

The sum of square divided by the corresponding degree of freedom gives the respective mean sum of square or variance.

$$\text{Mean sum of square due to treatment (MST)} = \frac{\text{SST}}{t - 1}$$

$$\text{Mean sum of square due to block (MSB)} = \frac{\text{SSB}}{r - 1}$$

$$\text{Mean sum of square due to error (MSE)} = \frac{\text{SSE}}{(t - 1)(r - 1)}$$

**ANOVA table**

S.V.	d.f.	S.S.	M.S.	$F_{\text{cal}}$	$F_{\text{tab}}$
Treatment	$t - 1$	SST	$MST = SST / (t - 1)$	$F_T = MST / MSE$	$F_{\alpha / (t-1), (t-1)(r-1)}$
Block	$r - 1$	SSB	$MSB = SSB / (r - 1)$	$F_B = MSB / MSE$	$F_{\alpha / (r-1), (t-1)(r-1)}$
Error	$(r-1)(t-1)$	SSE	$MSE = SSE / (t - 1)(r - 1)$		
Total	$rt - 1$	TSS			

**Decision:**

Reject  $H_0$  at  $\alpha$  % level of significance if  $F_T > F_{\alpha / (t-1), (t-1)(r-1)}$ , accept otherwise.

Reject  $H_0$  at  $\alpha$  % level of significance if  $F_B > F_{\alpha / (r-1), (t-1)(r-1)}$ , accept otherwise.

**Relation to calculate TSS, SST, SSB and SSE:**

$$\text{TSS} = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{G^2}{N}$$

$$= \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \text{C.F.}$$

$$\text{SST} = r \sum_{i=1}^t (y_{i.} - \bar{y}_{..})^2$$

$$= \frac{\sum_{i=1}^t T_i^2}{r} - \text{C.F.}$$

$$\text{SSB} = t \sum_{j=1}^r (\bar{y}_{.j} - \bar{y}_{..})^2$$

$$= \frac{\sum T_j^2}{r} - \text{C.F.}$$

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

### Advantages of RBD

- (i) RBD provides the better result than CRD.
- (ii) There is no restriction on the number of treatment or replication. But at least two replication is necessary.
- (iii) The statistical analysis remains simple if some observations are missing.

### Disadvantage of RBD

It is suitable for only small number of treatments.

**Example 3:** Carry out ANOVA of following design.

X	7	Y	8	Z	12	X	8
Y	10	X	9	X	6	Z	6
Z	8	Z	11	Y	9	Y	5

**Solution:**

Here, treatments are replicated along rowwise as only. Hence this is the case of RBD.

Problem to test

$H_{0T}$ : There is no significant difference between treatments.

$H_{1R}$ : There is significant difference between treatments.

$H_{0B}$ : There is no significant difference between block

$H_{1B}$ : There is significant difference between block.

Treatments	Block				$T_i$
	I	II	III	IV	
X	7	9	6	8	30
Y	10	8	9	5	32
Z	8	11	12	6	37
$T_j$	25	28	27	19	99

$N = 3 \times 4 = 12$

$$G = \sum_i T_i = \sum_j T_j = 99$$

$$C.F. = \frac{G^2}{N} = \frac{(99)^2}{12} = 816.75$$

$$\sum_{i,j} y_{ij}^2 = 7^2 + 9^2 + 6^2 + 8^2 + 10^2 + 8^2 + 9^2 + 5^2 + 8^2 + 11^2 + 12^2 + 6^2 = 865$$

$$TSS = \sum_{i,j} y_{ij}^2 - C.F. = 865 - 816.75 = 48.25$$

$$SST = \frac{1}{r} \sum_i T_i^2 - C.F.$$

$$= \frac{1}{4} [30^2 + 32^2 + 37^2] - 816.75 = 823.25 - 816.75 = 6.5$$

$$SSB = \frac{1}{t} \sum_i T_j^2 - C.F.$$

$$= \frac{1}{3} [25^2 + 28^2 + 27^2 + 19^2] - 816.75 = 833 - 816.75 = 16.25$$

$$SSE = TSS - SST - SSB$$

$$= 48.25 - 6.5 - 16.25$$

$$= 25.5$$

ANOVA table

S.V.	d.f.	S.S.	M.S.	$F_{Cal}$	$F_{Tab}$
Treatments	2	6.5	3.25	$F_T = 0.764$	$F_{0.05(2,6)} = 5.14$
Blocks	3	16.25	5.416	$F_R = 1.274$	$F_{0.05(3,6)} = 4.76$
Error	6	25.5	4.25		
Total	11	48.25			

**Decision**

$F_T = 0.764 < F_{0.05(2,6)} = 5.14$ , accept  $H_0$  at 5% level of significance.  
 $F_B = 1.274 < F_{0.05(3,6)} = 4.76$ , accept  $H_0$  at 5% level of significance.

**Conclusion**

There is no significant difference between treatments and there is no significant difference between blocks.

**Example 4:** Let A, B, C and D are four different page replacement algorithms. The following table gives the running times of programs under each replacement algorithm in 5 different blocks and in each block 5 different programs were used.

Block I		Block II		Block III		Block IV		Block V	
A	32	B	33	D	30	A	35	C	36
B	34	C	34	C	35	C	32	D	29
C	31	A	34	B	36	B	37	A	37
D	29	D	26	A	33	D	28	B	35

Analyse the above result to test whether there is significant difference between yields of four varieties.

**Problem to test**

$H_0$ : There is no significant difference between algorithms.

$H_1$ : There is significant difference between algorithms.

Here,

Treatment	Block				
	I	II	III	IV	V
A	32	34	33	35	37
B	34	33	36	37	35
C	31	34	35	32	36
D	29	26	30	28	29

Changing origin to 32 ( $u_{ij} = y_{ij} - 32$ )

Treatment	Block					$T_i$
	I	II	III	IV	V	
A	0	2	1	3	5	11
B	2	1	4	5	3	15
C	-1	2	3	0	4	8
D	-3	-6	-2	-4	-3	-18
$T_j$	-2	-1	6	4	9	16

$$\text{Now, } N = t \times r = 4 \times 5 = 20$$

$$G = \sum_i T_{i\cdot} = \sum_j T_{\cdot j} = 16$$

$$C.F. = \frac{G^2}{N} = \frac{(16)^2}{20} = 12.8$$

$$\sum_{i,j} u_{ij}^2 = (0 + 4 + 1 + 9 + 25 + 4 + 1 + 16 + 25 + 9 + 1 + 4 + 9 + 0 + 16 + 9 + 36 + 4 + 16 + 9) = 198$$

$$TSS = \sum_i \sum_j u_{ij}^2 - C.F.$$

$$= 198 - 12.8 = 185.2$$

$$SST = \frac{1}{r} \sum_i T_{i\cdot}^2 - C.F.$$

$$= \frac{1}{5} [(11)^2 + (15)^2 + (8)^2 + (-18)^2] - 12.8$$

$$= 146.8 - 12.8 = 134$$

$$SSB = \frac{1}{t} \sum_j T_{\cdot j}^2 - C.F.$$

$$= \frac{1}{4} [(-2)^2 + (-1)^2 + (6)^2 + (4)^2] - 12.8$$

$$= 34.5 - 12.8 = 21.7$$

$$SSE = TSS - SST - SSB$$

$$= 185.2 - 134 - 21.7$$

$$= 29.5$$

ANOVA table,

Source of variation	d.f.	S.S.	M.S.	F <sub>Cal</sub>	F <sub>Tab</sub>
Between treatment	3	134	46.66	18.982	F <sub>0.05(3,12)</sub> = 3.49
Between block	4	21.7	5.425	2.207	F <sub>0.05(4,12)</sub> = 3.26
Error	12	29.5	2.458		
Total	19	185.2			

Decision

$F_T = 18.982 > F_{0.05(3,12)} = 3.49$ , reject H<sub>0T</sub> at 5% level of significance.

Conclusion

There is significant difference between running times of four algorithms.

**Example 5:** Calculate the minimum number of replication so that an observed difference of 10% of the mean will be taken as significant at 5% level, the coefficient of variation (c.v.) of the plot values being 12%.

**Solution:**

Let  $r$  be the replication and  $\mu$  be the mean, then

$$C.V. = \frac{\sigma}{\mu} \times 100\%$$

$$\text{or } 12\% = \frac{\sigma}{\mu} \times 100\%$$

$$\text{or } \sigma = 0.12\mu \quad \dots\dots (*)$$

$$\text{Also, } \bar{x}_1 - \bar{x}_2 = 10\% \text{ of } \mu$$

$$\text{or } \bar{x}_1 - \bar{x}_2 = 0.1\mu \quad \dots\dots (**)$$

Now,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{2}{r}}} = \frac{0.1\mu}{0.12\mu \sqrt{\frac{2}{r}}} = \frac{5}{6} \sqrt{\frac{r}{2}}$$

For  $t$  to be significant at 5% level of significance, we have  $|t| > 1.96$

$$\text{or } \frac{5}{6} \sqrt{\frac{r}{2}} > 1.96$$

$$\text{or } r > (1.96)^2 \times (1.2)^2 \times 2$$

$$\text{or } r > 11.1$$

Hence required no of replication is 12.

#### Efficiency of RBD relative to CRD

The mechanism of the precision of RBD as compared to CRD is called efficiency of RBD relative to CRD.

Let us consider design having  $t$  treatments with  $r$  replication each.

If we perform RBD then,

The mathematical model of RBD is  $y_{ij} = \mu + \tau_i + \beta_j + e_{ij}; i = 1, 2, 3, \dots, t, j = 1, 2, 3, \dots, r$

If we perform CRD then,

The mathematical model of CRD is  $y_{ij} = \mu + \tau_i + e_{ij}; i = 1, 2, 3, \dots, t, j = 1, 2, 3, \dots, r$

Now the efficiency of RBD relative to CRD is given by

$$\text{Precision of RBD / Precision of CRD} = \frac{1}{MSE} / \frac{1}{MSE'} = \frac{MSE'}{MSE} = \frac{\sigma_e'^2}{\sigma_e^2} = \frac{r(t-1) MSE + (r-1) MSB}{(rt-1) MSE}$$

If  $\frac{\sigma_e'^2}{\sigma_e^2} < 1$  then RBD is less efficient than CRD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} > 1$  then RBD is more efficient than CRD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} = 1$  then RBD and CRD are equally effective.

**Example 6:** From the following ANOVA table of RBD, determine its efficiency with respect to CRD.

S.V.	d.f.	S.S.	M.S.S.
Between treatments	5	750	150
Between blocks	3	180	60
Error	15	200	13.33
Total	23	1130	

**Solution:**

Here,  $t = 6$ ,  $r = 4$ ,  $MSE = 13.33$ ,  $MST = 150$ ,  $MSB = 60$

Now, efficiency of RBD with respect to CRD is

$$\frac{s_e'^2}{s_e^2} = \frac{r(t-1) MSE + (r-1) MSB}{(rt-1) MSE} = \frac{4 \times 5 \times 13.33 + 3 \times 60}{23 \times 13.33} = 1.456$$

RBD is 45.6% efficient than CRD.

### Missing plot for RBD

Let us consider an RBD involving  $t$  treatment with  $r$  replication each. Let one of the observation say  $x$  receiving  $i^{th}$  treatment in the  $j^{th}$  block is missing.

Let,  $G'$  = total of all known  $rt-1$  values.

$T'$  = total of all known values of  $i^{th}$  treatment.

$B'$  = total of all known values of  $j^{th}$  block.

	$B_1$	$B_2$	$B_3 B_j$	$B_r$		$T_i$		
$T_1$	$y_{11}$	$y_{12}$	$y_{13}$	$\dots$	$y_{1j}$	$\dots$	$y_{1r}$	$T_1$
$T_2$	$y_{21}$	$y_{22}$	$y_{23}$	$\dots$	$y_{2j}$	$\dots$	$y_{2r}$	$T_2$
$T_3$	$y_{31}$	$y_{32}$	$y_{33}$	$\dots$	$y_{3j}$	$\dots$	$y_{3r}$	$T_3$
:	:	:	:		:		:	
$T_i$	$y_{i1}$	$y_{i2}$	$y_{i3}$	$\dots$	$x$	$\dots$	$y_{ir}$	$T_i = T' + x$
:	:	:	:		:		:	
:	:	:	:		:		:	
:	:	:	:		:		:	
$T_t$	$y_{t1}$	$y_{t2}$	$y_{t3}$	$\dots$	$y_{tj}$	$\dots$	$y_{tr}$	$T_t$
$T_j$	$T_{.1}$	$T_{.2}$	$T_{.3}$	$T_{.j} = B' + x$		$T_{.r}$	$G' + x$	

$$x = \frac{T't + B'r - G'}{(t-1)(r-1)}$$

Substitute the value of  $x$  in place of missing value and carry out analysis as usual except that one degree of freedom is subtracted from total and consequently from error. Because of the change in level of degree of freedom we obtain an upward bias in SST. Hence to get better result subtract an adjustment factor from SST.

$$\text{Adjustment factor (k)} = \frac{(B' + T't - G')^2}{t(t-1)(r-1)^2}$$

Adjusted SST ( $SST_A$ ) =  $SST - k$ .

**Example 7:** The table given below represents the yield of 3 varieties in 4 replicate experiment for which one observation is missing. Estimate the missing value and analyse the data.

A	18.1	B	?	A	15.2	C	13.2
C	16.0	A	12.1	B	17.5	A	16.6
B	16.3	C	13.4	C	16.3	B	18.1

**Solution:**

Treatment	Block				$T_i$
	I	II	III	IV	
A	18.1	12.1	15.2	16.6	62
B	16.3	? (x)	17.5	18.1	51.9+x
C	16.0	13.4	16.3	13.2	58.9
$T_j$	50.4	25.5+x	49.0	47.9	172.8+x

Here  $T' = 51.9$ ,  $B' = 25.5$ ,  $G' = 172.8$ ,  $t = 3$ ,  $r = 4$

$$\text{Now, missing value } x = \frac{T't + B'r - G'}{(t-1)(r-1)} = \frac{51.9 \times 3 + 25.5 \times 4 - 172.8}{2 \times 3} = 14.15$$

Problem to test

$H_{0T}$  : There is no significant difference between treatments.

$H_{1T}$  : There is significant difference between treatments.

$H_{0B}$  : There is no significant difference between blocks.

$H_{1B}$  : There is significant difference between blocks.

$$G = G' + x = 172.8 + 14.15 = 186.95, N = 4 \times 3 = 12$$

$$C.F. = \frac{G^2}{N} = \frac{(186.95)^2}{12} = 2912.525$$

$$\sum_i \sum_j y_{ij}^2 = (18.1)^2 + (12.1)^2 + (15.2)^2 + (16.6)^2 + (16.3)^2 + (14.15)^2 + (17.5)^2 + (18.1)^2 + 16^2 + (13.4)^2 + (16.3)^2 + (13.2)^2 = 2955.88$$

$$\begin{aligned} TSS &= \sum_i \sum_j y_{ij}^2 - C.F. \\ &= 2955.88 - 2912.525 = 43.354 \end{aligned}$$

$$\begin{aligned} SST &= \frac{1}{r} \sum_i T_i^2 - C.F. \\ &= \frac{1}{4} \{62^2 + 66.05^2 + 58.9^2\} - 2912.525 = 2918.953 - 2912.525 = 6.428 \end{aligned}$$

$$\begin{aligned} SSB &= \frac{1}{t} \sum_j T_{\cdot j}^2 - C.F. \\ &= \frac{1}{3} \{50.4^2 + 39.65^2 + 49^2 + 47.9^2\} - 2912.525 = 2935.897 - 2912.525 \\ &= 23.372 \end{aligned}$$

$$\text{Adjustment factor (k)} = \frac{(B' + tT' - G')^2}{t(t-1)(r-1)^2} = \frac{(25.5 + 3 \times 51.9 - 172.8)^2}{3 \times 2 \times 9} = 1.306$$

$$\begin{aligned} SST_A &= SST - k \\ &= 6.428 - 1.306 = 5.121 \end{aligned}$$

$$\begin{aligned} SSE &= TSS - SST_A - SSB \\ &= 43.354 - 5.121 - 23.372 \\ &= 14.861 \end{aligned}$$

S.V.	d.f.	S.S.	M.S.S	F <sub>Cal</sub>	F <sub>Tab</sub>
Treatments	2	5.121	2.56	0.861	$F_{0.05(2,5)} = 5.79$
Blocks	3	23.372	7.79	2.62	$F_{0.05(3,5)} = 5.41$
Error	5	14.861	2.972		
Total	10	43.354			

### Decision

$F_T = 0.861 < F_{0.05(2,5)} = 5.79$ , accept  $H_0T$  at 5% level of significance.

$F_B = 2.62 < F_{0.05(3,5)} = 5.41$ , accept  $H_0B$  at 5% level of significance.

### Conclusion

There is no significant difference between treatments.

There is no significant difference between blocks.

## Latin Square Design (LSD)

When the experimental material is not homogeneous the LSD is better than RBD. In RBD local control is used according to one way grouping i.e. according to blocks but in LSD local control is used according to two way grouping i.e. rows and columns. Hence it is used when two sources of errors are to be controlled simultaneously. In this design number of treatments are equal to the number of replication and the treatments are allocated in such a way that each of

the treatment occurs once and only once in each row and column. In this design Latin alphabet are used to denote the treatments, and shape is square due to equal number of treatments and replication so called Latin square design. It is based upon the all principles of design namely replication, randomization and local control.

### Lay out

Let us consider  $m$  treatments with  $m$  replication each so that there are  $N = m^2$  experimental unit. Let us divide the experimental material into  $m^2$  experimental units arranged in square so that each row as well as column contains  $m$  units. In this design none of treatments are replicated along row wise or column wise. In this case we study the variation between treatments, the variation between rows and variation between columns. It has only  $m^2$  experimental unit but studies variation of three factors i.e. rows, columns and treatments. Hence it is the case of incomplete three way ANOVA. For complete three way ANOVA we need  $m^3$  experimental unit.

Let us consider  $t = 4(A, B, C, D)$  then  $4 \times 4$  LSD is as shown below.

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

### Mathematical model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}; i = 1, 2, 3 \dots m; j = 1, 2, 3 \dots m; k = 1, 2, 3 \dots m$$

Where,

$y_{ijk}$  =  $i^{th}$  row and  $j^{th}$  column receiving  $k^{th}$  treatment.

$\mu$  = constant effect

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

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

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

$e_{ijk}$  = error due to chance

### Assumptions

- All the observations are independent.
- All the observations should be drawn from normal population having constant variance.
- All the treatments should be homogeneous as far as possible.
- Various treatments and environmental effect are additive in nature.
- $e_{ijk}$  are i.i.d.  $N(0, \sigma_e^2)$

### Statistical analysis:

In the model  $y_{ij} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}$ , parameters  $\mu, \alpha_i, \beta_j$  and  $\tau_k$  are determined by using the principle of least square by minimizing error(residual) sum of square.

Total sum of square (TSS) = Sum of square due to row(SSR) + Sum of square due to column (SSC) + Sum of square due to treatment (SST) + Sum of square due to error (SSE)

Degree of freedom (d.f) for various sum of square:

Degree of freedom for total sum of square =  $m^2 - 1 = N - 1$

Degree of freedom for sum of square due to row =  $m-1$

Degree of freedom for sum of square due to column =  $m-1$

Degree of freedom for sum of square due to treatment =  $m-1$

Mean Sum of Square (MSS)

The sum of square divided by the corresponding degree of freedom gives the respective mean sum of square or variance.

$$\text{Mean sum of square due to row (MSR)} = \frac{\text{SSR}}{m-1}$$

$$\text{Mean sum of square due to column (MSC)} = \frac{\text{SSC}}{m-1}$$

$$\text{Mean sum of square due to treatment (MST)} = \frac{\text{SST}}{m-1}$$

$$\text{Mean sum of square due to error (MSE)} = \frac{\text{SSE}}{(m-1)(m-2)}$$

**ANOVA table**

S.V.	d.f.	S.S.	M.S.	$F_{\text{cal}}$	$F_{\text{tab}}$
Row	$m-1$	SSR	$\text{MSR} = \frac{\text{SSR}}{m-1}$	$F_R = \text{MSR}/\text{MSE}$	$F_{\alpha((m-1),(m-1)(m-2))}$
Column	$m-1$	SSC	$\text{MSC} = \frac{\text{SSC}}{m-1}$	$F_C = \text{MSC}/\text{MSE}$	$F_{\alpha((m-1),(m-1)(m-2))}$
Treatment	$m-1$	SST	$\text{MST} = \frac{\text{SST}}{m-1}$	$F_T = \text{MST}/\text{MSE}$	$F_{\alpha((m-1),(m-1)(m-2))}$
Error	$(m-1)(m-2)$	SSE	$\text{MSE} = \frac{\text{SSC}}{(m-1)(m-2)}$		
Total	$m^2 - 1$	TSS			

Decision:

Reject  $H_0R$  at  $\alpha\%$  level of significance if  $F_R > F_{\alpha((m-1),(m-1)(m-2))}$ , accept otherwise.

Reject  $H_0C$  at  $\alpha\%$  level of significance if  $F_C > F_{\alpha((m-1),(m-1)(m-2))}$ , accept otherwise.

Reject  $H_0T$  at  $\alpha\%$  level of significance if  $F_T > F_{\alpha((m-1),(m-1)(m-2))}$ , accept otherwise.

Relation to calculate TSS, SSR, SSC, SST and SSE:

$$\begin{aligned} \text{TSS} &= \sum_{(i,j,k)} (y_{ijk} - \bar{y}_{...})^2 = \sum_{(i,j,k)} y_{ijk}^2 - \frac{G^2}{N} = \sum_{(i,j,k)} y_{ijk}^2 - \text{C.F.} \end{aligned}$$

$$\begin{aligned} \text{SSR} &= m \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \frac{\sum_i T_{i..}^2}{m} - \text{CF} \end{aligned}$$

$$\begin{aligned} \text{SSC} &= m \sum_i (\bar{y}_{.j.} - \bar{y}_{...})^2 = \frac{\sum_i T_{.j..}^2}{m} - \text{CF} \end{aligned}$$

$$\begin{aligned} \text{SST} &= m \sum_i (\bar{y}_{..k} - \bar{y}_{...})^2 = \frac{\sum_i T_{..k..}^2}{m} - \text{CF} \end{aligned}$$

$$\begin{aligned} \text{SSE} &= TSS - SSR - SSC - SST \end{aligned}$$

**Advantages of LSD**

- i) Due to the use of two way grouping of controls more variation than CRD and RBD.
- ii) It is incomplete three way layout. Its advantage over complete three way layout is that instead of  $m^3$  experimental units only  $m^2$  units are needed.
- iii) The statistical analysis remains simple if some observations are missing.

**Disadvantages of LSD**

- i) The assumption of factors are independent is not always true.
- ii) It is suitable for treatments 5 to 10.
- iii) It is not easy in the field layout.

**Example 8:** The following is the  $5 \times 5$  Latin square design for data taken from a manorial experiment with sugarcane. The five treatments were A = no manure; B = an inorganic manure;

C, D and E = three levels of farm yard manure.

Plan and yield of sugarcane (in a suitable unit) per plot.

Row	Column				
	I	II	III	IV	V
I	A 52.5	E 46.3	D 44.1	C 48.1	B 40.9
II	D 44.2	B 42.9	A 51.3	E 49.3	C 32.6
III	B 49.1	A 47.3	C 38.1	D 41.0	E 47.2
IV	C 43.2	D 42.5	E 67.2	B 55.1	A 45.3
V	E 47.0	C 43.2	B 46.7	A 46.0	D 43.2

Analyze the above data to find if there are any treatment effects.

**Solution:**

Problem to test

$H_{0T}$ : There is no significant difference between treatments.

$H_{1T}$ : There is significant difference between treatments.

Row	Column					
	I	II	III	IV	V	T <sub>ti..</sub>
I	A 52.5	E 46.3	D 44.1	C 48.1	B 40.9	231.9
II	D 44.2	B 42.9	A 51.3	E 49.3	C 32.6	220.3
III	B 49.1	A 47.3	C 38.1	D 41.0	E 47.2	222.7
IV	C 43.2	D 42.5	E 67.2	B 55.1	A 45.3	253.3
V	E 47.0	C 43.2	B 46.7	A 46.0	D 43.2	226.1
T <sub>j..</sub>	236.0	222.2	247.4	239.5	209.2	1154.3

$$\begin{aligned}
 T_{..A} &= 52.5 + 51.3 + 47.3 + 45.3 + 46.0 = 242.4 \\
 T_{..B} &= 40.9 + 42.9 + 49.1 + 55.1 + 46.7 = 234.7 \\
 T_{..C} &= 48.1 + 32.6 + 38.1 + 43.2 + 43.2 = 205.2 \\
 T_{..D} &= 44.1 + 44.2 + 41.0 + 42.5 + 43.2 = 215.0 \\
 T_{..E} &= 46.3 + 49.3 + 47.2 + 67.2 + 47.0 = 257.0 \\
 G &= 1154.3, N = 25
 \end{aligned}$$

$$C.F. = \frac{G^2}{N} = \frac{(1154.3)^2}{25} = 53296.33$$

$$\sum_{(i,j,k)} y_{ijk}^2 = (52.5)^2 + (46.3)^2 + (44.1)^2 + (48.1)^2 + (40.9)^2 + \dots + (43.2)^2 = 54273.51$$

$$TSS = \sum_{(i,j,k)} y_{ijk}^2 - C.F. = 54273.51 - 53296.33 = 977.18$$

$$\begin{aligned}
 SSR &= \frac{\sum_i T_{i..}^2}{m} - C.F. = \frac{1}{5} \{ (231.9)^2 + (220.3)^2 + (222.7)^2 + (253.3)^2 + (226.1)^2 \} - 53296.33 \\
 &= 53437.41 - 53296.33 = 141.08
 \end{aligned}$$

$$\begin{aligned}
 SSC &= \frac{\sum_j T_{j..}^2}{m} - C.F. = \frac{1}{5} \{ (236)^2 + (222.2)^2 + (247.4)^2 + (239.5)^2 + (209.2)^2 \} - 53296.33 \\
 &= 53480.09 - 53296.33 \\
 &= 183.76
 \end{aligned}$$

$$\begin{aligned}
 SST &= \sum_k \frac{T_{..k}^2}{m} - C.F. = \frac{1}{5} \{ (242.4)^2 + (234.7)^2 + (205.2)^2 + (215)^2 + (257)^2 \} - 53296.33 \\
 &= 53644.57 - 53296.33 \\
 &= 348.24
 \end{aligned}$$

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

$$= 977.18 - 141.08 - 183.76 - 348.24 = 304.1$$

ANOVA table

S.V.	d.f.	S.S.	M.S.	F <sub>Cal</sub>	F <sub>Tab</sub>
Row	4	141.08	35.27		
Column	4	183.76	45.94		
Treatment	4	348.24	87.06	F <sub>T</sub> = 3.436	F <sub>0.05(4,12)</sub> = 3.26
Error	12	304.1	25.34		
Total	24	977.18			
Decision					

$F_T = 3.436 > F_{0.05(4,12)} = 3.26$ , reject  $H_0 T$  at 5% level of significance.

Conclusion

There are treatment effects.

**Example 9:** The layout and yield of four treatments in a  $4 \times 4$  experiment is shown in the following table. Analyse the data.

D 20	B 17	A 20	C 19
B 21	A 18	C 18	D 17
A 18	C 21	D 17	B 17
C 20	D 19	B 17	A 18

**Solution:**

Problem to test

$H_{0R}$ : There is no significant difference between rows

$H_{1R}$ : There is significant difference between rows.

$H_{0C}$ : There is no significant difference between columns

$H_{1C}$ : There is significant difference between columns.

$H_{0T}$ : There is no significant difference between treatments

$H_{1T}$ : there is significant difference between treatments.

Now, changing origin to 19 ( $u_{ijk} = y_{ijk} - 19$ )

					T <sub>i..</sub>
	D 1	B -2	A 1	C 0	0
	B 2	A -1	C -1	D -2	-2
	A -1	C 2	D -2	B -2	-3
	C 1	D 0	B -2	A -1	-2
T <sub>.j..</sub>	3	-1	-4	-5	-7

$$G = -7, N = 4^2 = 16$$

$$C.F. = \frac{G^2}{N} = \frac{(-7)^2}{16} = 3.0625$$

$$\sum_{(i,j,k)} u_{ijk}^2 = 1 + 4 + 1 + 0 + 4 + 1 + 1 + 4 + 1 + 4 + 4 + 4 + 1 + 0 + 4 + 1 = 35$$

$$TSS = \sum_{(i,j,k)} u_{ijk}^2 - C.F. = 35 - 3.0625 = 31.9375$$

$$SSR = \frac{1}{4} \sum_i T_{i..}^2 - C.F. = \frac{1}{4} \{(0)^2 + (-2)^2 + (-3)^2 + (-2)^2\} - 3.0625 = 3.75 - 3.0625 = 0.6875$$

$$SSC = \frac{1}{4} \sum_j T_{.j..}^2 - C.F. = \frac{1}{4} \{(3)^2 + (-1)^2 + (-4)^2 + (-5)^2\} - 3.0625 = 12.75 - 3.0625 = 9.6875$$

$$T_{..A} = 1 - 1 - 1 - 1 = -2$$

$$T_{..B} = -2 + 2 - 2 - 2 = -4$$

$$T_{..C} = 0 - 1 + 2 + 1 = 2$$

$$T_{..D} = 1 - 2 - 2 + 0 = -3$$

$$SST = \frac{1}{4} \sum_k T_{..k}^2 - C.F. = \frac{1}{4} [(-2)^2 + (-4)^2 + (2)^2 + (-3)^2] - 3.0625 = 8.25 - 3.0625 = 5.1875$$

$$SSE = TSS - SSR - SSC - SST = 31.9375 - 0.6875 - 9.6875 - 5.1875 = 16.375$$

S.V.	d.f.	S.S.	M.S.	F <sub>Cal</sub>	F <sub>Tab</sub>
Row	3	0.6875	0.2291	0.0839	F <sub>0.05(3,6)</sub> = 4.76
Column	3	9.6875	3.2291	1.1832	F <sub>0.05(3,6)</sub> = 4.76
Treatment	3	5.1875	1.7291	0.6335	F <sub>0.05(3,6)</sub> = 4.76
Error	6	16.375	2.7291		
Total	15	31.9375			

**Decision**

F<sub>R</sub> = 0.0839 < F<sub>0.05(3,6)</sub> = 4.76, accept H<sub>0R</sub> at 5% level of significance.

F<sub>C</sub> = 1.1832 < F<sub>0.05(3,6)</sub> = 4.76, accept H<sub>0C</sub> at 5% level of significance.

F<sub>T</sub> = 0.6335 < F<sub>0.05(3,6)</sub> = 4.76, accept H<sub>0T</sub> at 5% level of significance.

**Conclusion**

There is no significant difference between rows, there is no significant difference between columns and there is no significant difference between treatments.

**Efficiency of LSD Relative to CRD**

The mechanism of the precision of LSD as compared to CRD is called efficiency of LSD relative to CRD.

If we perform LSD then,

The mathematical model of LSD is  $y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}$ ;  $i = 1, 2, 3 \dots m$ ;  $j = 1, 2, 3 \dots m$ ;  $k = 1, 2, 3 \dots m$

If we perform CRD then,

The mathematical model of CRD is  $y_{ij} = \mu + \tau_i + e_{ij}$ ;  $i = 1, 2, 3 \dots m$ ,  $j = 1, 2, 3 \dots m$

Now the efficiency of LSD relative to CRD is given by

$$\begin{aligned} \text{Precision of LSD / Precision of CRD} &= \frac{1}{MSE} / \frac{1}{MSE'} = \frac{MSE'}{MSE} = \frac{\sigma_e'^2}{\sigma_e^2} \\ &= \frac{(m-1) MSE + MSR + MSC}{(m+1) MSE} \end{aligned}$$

If  $\frac{\sigma_e'^2}{\sigma_e^2} < 1$  then LSD is less efficient than CRD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} > 1$  then LSD is more efficient than CRD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} = 1$  then LSD and CRD are equally effective.

**Example 10:** From the following ANOVA table of  $4 \times 4$  LSD determine its efficiency with respect to CRD.

S.V.	d.f.	S.S.	M.S.S.
Rows	3	2.133	0.711
Columns	3	2.203	0.734
Treatments	3	10.663	3.554
Error	6	7.059	1.177
Total	15	22.058	

**Solution:**

$$m = 4, \text{MSR} = 0.711, \text{MSC} = 0.734, \text{MST} = 3.554, \text{MSE} = 1.177$$

Now,

Efficiency of LSD with respect to CRD is

$$\frac{\sigma_e'^2}{\sigma_e^2} = \frac{(m-1) \text{ MSE} + \text{MSR} + \text{MSC}}{(m+1) \text{ MSE}} = \frac{3 \times 1.177 + 0.711 + 0.734}{5 \times 1.177} = 0.845 = 84.5\%$$

LSD is 15.5% less efficient than CRD.

#### Efficiency of LSD relative to RBD

The mechanism of the precision of LSD as compared to RBD is called efficiency of LSD relative to RBD.

If we perform LSD then,

The mathematical model of LSD is  $y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}; i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, m; k = 1, 2, 3, \dots, m$

If we perform RBD then,

The mathematical model of RBD is  $y_{ij} = \mu + \tau_i + \beta_j + e_{ij}; i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, m$

(i) When row is taken as block:

$$\frac{\sigma_e'^2}{\sigma_e^2} = \frac{(m-1) \text{ MSE} + \text{MSC}}{m \text{ MSE}}$$

If  $\frac{\sigma_e'^2}{\sigma_e^2} < 1$  then LSD is less efficient than RBD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} > 1$  then LSD is more efficient than RBD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} = 1$  then LSD and RBD are equally effective.

When column is taken as block:

$$\frac{\sigma_e'^2}{\sigma_e^2} = \frac{(m-1)MSE + MSR}{mMSE}$$

If  $\frac{\sigma_e'^2}{\sigma_e^2} < 1$  then LSD is less efficient than RBD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} > 1$  then LSD is more efficient than RBD.

If  $\frac{\sigma_e'^2}{\sigma_e^2} = 1$  then LSD and RBD are equally effective.

**Example 11:** From the following ANOVA table of 4x4 LSD determine it's efficiency with respect to RBD.

S.V.	d.f.	S.S.	M.S.S.
Rows	3	2.133	0.711
Columns	3	2.203	0.734
Treatments	3	10.663	3.554
Error	6	7.059	1.177
Total	15	22.058	

**Solution:**

$m = 4$ ,  $MSR = 0.711$ ,  $MSC = 0.734$ ,  $MST = 3.554$ ,  $MSE = 1.177$

When row is taken as block

$$\frac{\sigma_e'^2}{\sigma_e^2} = \frac{(m-1)MSE + MSC}{mMSE} = \frac{3 \times 1.177 + 0.734}{4 \times 1.177} = 0.905 = 90.5\%$$

Hence, LSD is 9.5% less efficient than RBD when row is taken as block.

When column is taken as block

$$\frac{\sigma_e'^2}{\sigma_e^2} = \frac{(m-1)MSE + MSR}{mMSE} = \frac{3 \times 1.177 + 0.711}{4 \times 1.177} = 0.901 = 90.1\%$$

Hence LSD is 9.9% less efficient than RBD when column is taken as block.

#### Missing plot for LSD

Let us consider a  $m \times m$  LSD. Let one of the observation say  $x$  occurring in  $i^{th}$  row,  $j^{th}$  column and  $k^{th}$  treatment is missing.

Let,  $G' = \text{total of all known } m^2 - 1 \text{ values.}$

$R' = \text{total of all known values of } i^{th} \text{ row}$

$C' = \text{total of all known values of } j^{th} \text{ column}$

$T' = \text{total of all known values of } k^{th} \text{ treatment.}$

	Column						R <sub>i..</sub>	
Row	y <sub>111</sub>	y <sub>122</sub>	y <sub>133</sub>	.....	y <sub>1j4</sub>	.....	y <sub>1mm</sub>	R <sub>1..</sub>
	y <sub>212</sub>	y <sub>223</sub>	y <sub>234</sub>	.....	y <sub>2j5</sub>	.....	y <sub>2m1</sub>	R <sub>2..</sub>
	y <sub>313</sub>	y <sub>324</sub>	y <sub>335</sub>	.....	y <sub>3j6</sub>	.....	y <sub>3m2</sub>	R <sub>3..</sub>
	:	:	:	:	:	:	:	
	y <sub>i14</sub>	y <sub>i25</sub>	y <sub>i36</sub>	.....	x	.....	y <sub>im3</sub>	R <sub>i..</sub> = R' + x
	:	:	:	:	:	:	:	
	:	:	:	:	:	:	:	
	:	:	:	:	:	:	:	
	y <sub>m1m</sub>	y <sub>m21</sub>	y <sub>m32</sub>	.....	y <sub>mj3</sub>	.....	y <sub>mm6</sub>	R <sub>m..</sub>
C <sub>.i.</sub>	C <sub>.1.</sub>	C <sub>.2.</sub>	C <sub>.3.</sub>	C <sub>.j.</sub> = C' + x		C <sub>.m.</sub>	G' + x	

$$x = \frac{m(R' + C' + T') - 2G'}{(m-1)(m-2)}$$

Substitute the value of x in place of missing value and carry out analysis as usual except that one degree of freedom is subtracted from total and consequently from error. Because of the change in level of degree of freedom we obtain an upward bias in SST. Hence to get better result subtract an adjustment factor from SST.

$$\text{Adjustment factor (k)} = \frac{(m-1)T' + R' + C' - G'}{(m-1)(m-2)^2}$$

$$\text{Adjusted SST (SST}_A\text{)} = \text{SST} - k.$$

**Example 12:** Determine the missing value and carryout ANOVA of following design.

D	20.1	B	19.4	C	30.6	A	7.9
C	17.5	A	10.4	D	21.2	B	19.1
A	(x)	D	18.1	B	24.6	C	25.2
B	25.1	C	30.4	A	10.2	D	28.0

**Solution:**

T <sub>i..</sub>	D	B	C	A	T <sub>.i.</sub>			
D	20.1	B	19.4	C	30.6	A	7.9	78
C	17.5	A	10.4	D	21.2	B	19.1	68.2
A	(x)	D	18.1	B	24.6	C	25.2	67.9+x
B	25.1	C	30.4	A	10.2	D	28.0	93.7
T <sub>.j.</sub>	62.7+x	78.3	86.6	80.2	307.8+x			

Here,

$$m = 4, R' = 67.9, C' = 62.7, T' = 10.4 + 10.2 + 7.9 = 28.5, G' = 307.8$$

$$\begin{aligned}\text{Missing value (x)} &= \frac{m(R' + C' + T') - 2G'}{(m-1)(m-2)} \\ &= \frac{4(67.9 + 62.7 + 28.5) - 2 \times 307.8}{3 \times 2} \\ &= \frac{636.4 - 615.6}{6} = 3.46\end{aligned}$$

**Problem to test**

$H_0:$  There is no significant difference between rows

$H_1:$  There is significant difference between rows.

$H_0:$  There is no significant difference between columns

$H_1:$  There is significant difference between columns.

$H_0:$  There is no significant difference between treatments

$H_1:$  there is significant difference between treatments.

Now,

$$G = G' + x = 307.8 + 3.46 = 311.26$$

$$N = m^2 = 4^2 = 16$$

$$C.F. = \frac{G^2}{N} = \frac{(311.26)^2}{16} = 6055.17$$

$$\begin{aligned}\sum_{i,j,k} y_{ijk}^2 &= 20.1^2 + 19.4^2 + 30.6^2 + 7.9^2 + 17.5^2 + 10.4^2 + 21.2^2 + 19.1^2 + 3.46^2 + 18.1^2 + 24.6^2 + \\ &\quad 25.2^2 + 25.1^2 + 30.4^2 + 10.2^2 + 28.0^2 \\ &= 7029.79\end{aligned}$$

$$\begin{aligned}TSS &= \sum_{i,j,k} y_{ijk}^2 - C.F. \\ &= 7029.79 - 6055.17 = 974.62\end{aligned}$$

$$\begin{aligned}SSR &= \frac{1}{m} \sum_i T_{i..}^2 - C.F. \\ &= \frac{1}{4} \{78^2 + 68.2^2 + 71.36^2 + 93.7^2\} - 6055.17 = 6151.79 - 6055.17 \\ &= 96.62\end{aligned}$$

$$\begin{aligned}SSC &= \frac{1}{m} \sum_j T_{..j}^2 - C.F. \\ &= \frac{1}{4} \{66.16^2 + 78.3^2 + 86.6^2 + 80.2^2\} - 6055.17 = 6109.9 - 6055.17\end{aligned}$$

$$\begin{aligned}T_{..A} &= 54.73 \\ &= 3.46 + 10.4 + 10.2 + 7.9 \\ &= 31.96\end{aligned}$$

$$\begin{aligned}
 T_{..B} &= 25.1 + 19.4 + 24.6 + 19.1 \\
 &= 88.2 \\
 T_{..C} &= 17.5 + 30.4 + 30.6 + 25.2 \\
 &= 103.7 \\
 T_{..D} &= 20.1 + 18.1 + 21.2 + 28 \\
 &= 87.4
 \end{aligned}$$

$$\begin{aligned}
 SST &= \frac{1}{m} \sum_i T_{..k^2} - C.F. \\
 &= \frac{1}{4} \{31.96^2 + 88.2^2 + 103.7^2 + 87.4^2\} - 6055.17 \\
 &= 6798.28 - 6055.17 \\
 &= 743.11
 \end{aligned}$$

$$\begin{aligned}
 \text{Adjustment factor (k)} &= \frac{\{(m-1)T' + R' + C' - G'\}^2}{\{(m-1)(m-2)\}^2} \\
 &= \frac{\{3 \times 28.5 + 67.9 + 62.7 - 307.8\}^2}{\{3 \times 2\}^2} = 233.58
 \end{aligned}$$

$$\begin{aligned}
 SST_A &= SST - k \\
 &= 743.11 - 233.58 = 509.53
 \end{aligned}$$

$$\begin{aligned}
 SSE &= TSS - SSR - SSC - SST_A \\
 &= 974.62 - 96.62 - 54.73 - 509.53 \\
 &= 313.94
 \end{aligned}$$

ANOVA table

S.V.	d.f.	S.S.	M.S.S.	F <sub>Cal</sub>	F <sub>Tab</sub>
Rows	3	96.62	32.2	0.513	F <sub>0.05(3,5)</sub> = 5.41
Columns	3	54.73	18.24	0.29	F <sub>0.05(3,5)</sub> = 5.41
Treatments	3	509.53	169.84	2.706	F <sub>0.05(3,5)</sub> = 5.41
Error	5	313.74	62.748		
Total	14				

**Decision**

$F_R = 0.513 < F_{0.05(3,5)} = 5.41$ , accept  $H_{0R}$  at 5% level of significance.

$F_C = 0.29 < F_{0.05(3,5)} = 5.41$ , accept  $H_{0C}$  at 5% level of significance.

$F_T = 2.706 < F_{0.05(3,5)} = 5.41$ , accept  $H_{0T}$  at 5% level of significance.

**Conclusion**

There is no significant difference between rows. There is no significant difference between columns. There is no significant difference between treatments.

**EXERCISE**

1. Describe basic principles of experimental design.
2. Explain the terms with examples; Experiment, Treatments, Experimental units, Blocks, Experimental error, Precision.
3. What do you mean by CRD? Write down its advantages and disadvantages.
4. What do you mean by RBD? Write down its advantages and disadvantages.
5. What do you mean by LSD? Write down its advantages and disadvantages.
6. Differentiate between CRD and RBD.
7. Differentiate between RBD and LSD.
8. What are the assumptions underlying the analysis of the results of LSD.
9. Clearly state the restrictions that are being imposed on the number of treatments and number of replications in CRD, RBD and LSD.
10. What is meant by relative efficiency? Give expression for efficiency of (i) LSD relative to RBD. (ii) LSD relative to CRD (iii) RBD relative to CRD.
11. Carry out ANOVA of following output of wheat per field obtained as a result of 3 varieties of wheat A, B and C.

A 10	B 5	A 20	C 15
B 6	A 15	C 11	B 10
C 22	B 12	C 18	A 16

Ans:  $F_T = 4.793$ , insig.

12. Carry out ANOVA for following design.

A 8	C 10	A 6	B 10
C 12	B 8	B 9	A 8
B 10	A 8	C 10	C 9

Also calculate the relative efficiency of the design with respect to CRD.

Ans:  $F_T = 7.97$  insig.,  $F_B = 1.6$ , insig., Efficiency = 1.163

13. Set up the analysis of variance for the following results of a design.

A 10	B 15	C 20
B 25	C 10	A 15
C 25	A 20	B 15

Also calculate the efficiency of the design over i) RBD ii) CRD.

Ans:  $F_R = 0.25$ ,  $F_C = 0.25$ ,  $F_T = 0.142$  insig. Efficiency = 0.625, 0.75, 0.75

14. The table given below are yields of 3 varieties in a 4 replicate experiment for which one observation is missing. Estimate the missing value and then analyse the data.

P 19	R 29	P 23	Q 33
Q 26	P ?	Q 27	R 26
R 21	Q 28	R 22	P 26

Ans: 25.3,  $F_T = 4.69$ ,  $F_B = 4.72$ , insig.

15. The table given below represents the yields of 4 varieties in a 4 replicate experiment for which one observation is missing. Estimate the missing value and then carry out the ANOVA

A 12	C 19	B 10	D 8
C 18	B 12	D 6	A ?
B 22	D 10	A 5	C 21
D 12	A 7	C 27	B 17

Ans: 16.67, FR = 0.26, FC = 0.559, FT = 3.32, insig.

16. Complete the following table for the analysis of variance of a design.

S.V.	d.f.	S.S.	M.S.S.	F
Blocks	4	26.8	?	?
Treatment	3	?	?	?
Error	?	?	2.5	
Total	?	85.3		

Ans: 12, 19, 28.5, 30, 6.7, 9.5, 2.68, 3.8

17. Fill in the blanks in the following analysis of variance table of a design.

Source of Variation	d.f.	S.S.	M.S.S.	F
Rows	?	72	?	2
Columns	?	?	36	?
Treatments	?	180	?	?
Error	6	?	12	
Total	?	?		

Ans: 3, 3, 3, 15, 108, 72, 432, 24, 60, 3, 5

18. Complete the following table for analysis of variance of a design.

Source of variation	Degree of freedom	Sum of squares	Mean square	F
Columns	5	?	?	?
Rows	?	4.2	?	?
Treatments	?	?	2.43	?
Error	?	?	0.65	
Total	?	39.65		

The columns as representing schools, the rows as classes, the treatments as methods of teaching and the observations as grades based on 100 points. Test the hypothesis that the treatment effects are equal to zero.

Ans: 5, 5, 20, 35, 10.3, 12.15, 13, 2.06, 0.84, 3.16, 1.29, 3.73, Reject H<sub>0</sub>

19. Consider the partially completed ANOVA table below. Complete the ANOVA table and answer the followings. What design was employed? How many treatments were compared? How many observations were analyzed? At 0.05 level of significance, can one conclude that the treatments have different effects? Why?

Source of variation	Sum of Square	Degree of freedom	Mean Square	F
Treatments	231.5	2	?	?
Blocks	?	7	?	?
Error	573.75	?	?	
Total	903.75	23		

**Ans:** 98.5, 14, 115.75, 14.07, 40.98, 2.82, 0.34, Accept H<sub>0</sub>

20. From the following ANOVA table of RBD, determine it's efficiency with respect to CRD.

Source	D.F.	S.S.	M.S.S.
Between Blocks	5	21.55	4.31
Between Treatments	3	15.66	5.22
Error	15	12.3	0.82
Total	23	49.51	

**Ans:** 1.925

21. From the following ANOVA table of LSD , determine it's efficiency i)with respect to CRD ii) with respect to RBD when columns are taken as blocks iii) with respect to RBD when rows are taken as blocks.

Source of variation	Degree of freedom	Sum of squares	Mean sum of squares
Rows	3	259.5375	86.4375
Columns	3	155.2725	51.7575
Treatments	3	1372.1225	457.3742
Error	6	156.3700	26.0616
Total	15	1943.0775	

**Ans:** 1.6605, 1.6605, 1.2464



## Using Software

### CRD

The yield of treatments in different plots are as shown in the following plots. Carry out analysis.

t <sub>4</sub> 1401	t <sub>3</sub> 2536	t <sub>3</sub> 2459	t <sub>1</sub> 2537	t <sub>3</sub> 2827	t <sub>1</sub> 2069
t <sub>2</sub> 2211	t <sub>1</sub> 1797	t <sub>4</sub> 1170	t <sub>4</sub> 1516	t <sub>4</sub> 2104	t <sub>3</sub> 2385
t <sub>2</sub> 3366	t <sub>1</sub> 2104	t <sub>2</sub> 2591	t <sub>3</sub> 2460	t <sub>4</sub> 1077	t <sub>2</sub> 2544

### Using excel

Data\ Data analysis\ Anova: Single Factor

t <sub>4</sub>	1401	t <sub>3</sub>	2536	t <sub>3</sub>	2459	t <sub>1</sub>	2537	t <sub>3</sub>	2827	t <sub>1</sub>	2069
t <sub>2</sub>	2211	t <sub>1</sub>	1797	t <sub>4</sub>	1170	t <sub>4</sub>	1516	t <sub>4</sub>	2104	t <sub>3</sub>	2385
t <sub>2</sub>	3366	t <sub>1</sub>	2104	t <sub>2</sub>	2591	t <sub>3</sub>	2460	t <sub>4</sub>	1077	t <sub>2</sub>	2544
t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	t <sub>4</sub>								
2537	2211	2536	1401								
2069	3366	2459	1170								
1797	2597	2827	1516								
2104	2544	2385	2104								
1077		2460									

A	B	C	D	E	F	G
15	Anova: Single Factor					
16						
17	SUMMARY					
18	Groups	Count	Sum	Average	Variance	
19	t <sub>1</sub>	5	9584	1916.8	290618.2	
20	t <sub>2</sub>	4	10718	2679.5	238647	
21	t <sub>3</sub>	5	12667	2533.4	29788.3	
22	t <sub>4</sub>	4	6191	1547.75	158217.6	
23						
24						
25	ANOVA					
26	Source of Variation	SS	df	MS	F	P-value
27	Between Groups	3567435	3	1189145	6.734041	0.004844
28	Within Groups	2472220	14	176587.1		3.343889
29						
30	Total	6039654	17			
31						

Problem to test

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_A:$  At least one of the  $\mu_i$  is different,  $i = 1, 2, 3, 4$  ( $1 = t_1, 2 = t_2, 3 = t_3, 4 = t_4$ )

A	B	C	D	E	F	G
31 $\alpha$	0.05					
32 Decision						
33 the null hypothesis $H_0$ is rejected						
34						
35						
36 it is significant						
37						

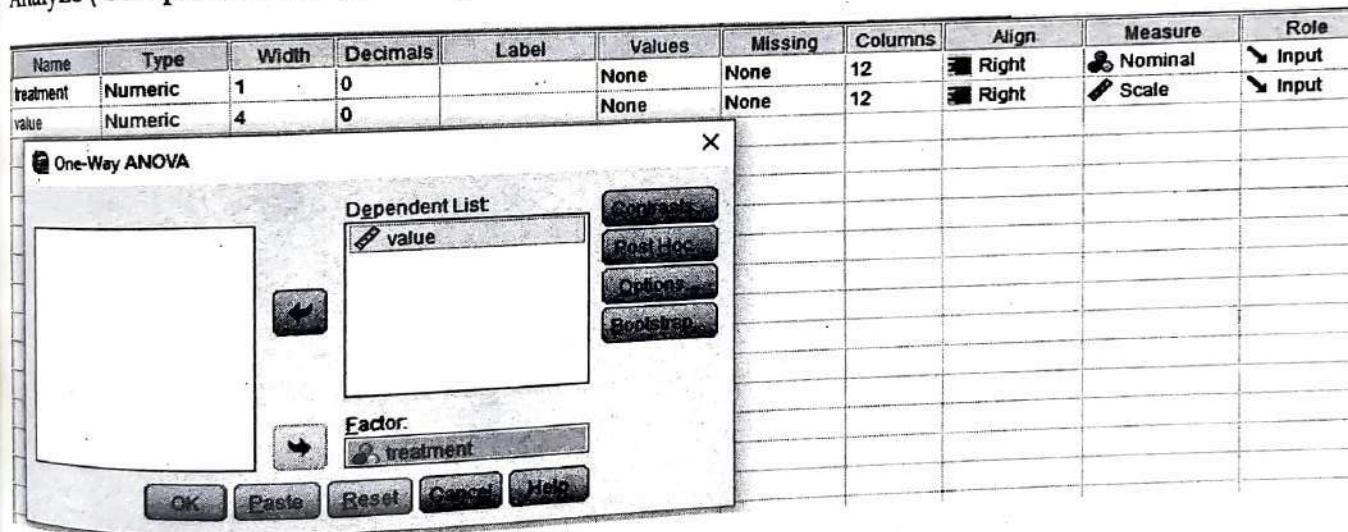
### Formula

=IF(E26<G26, "there is no reason to reject Null hypothesis  $H_0$ ", "the null hypothesis  $H_0$  is rejected")  
=IF(F26<B31, "it is significant", "it is not significant")

### Using SPSS

Create variable for treatment and its values

Analyze\Compare means\One way anova



Select LSD in Post HOC test

### ANOVA

value	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3561394.361	3	1187131.454	6.720	.005
Within Groups	2473236.750	14	176659.768		
Total	6034631.111	17			

## Multiple Comparisons

Dependent Variable: value

LSD

(I) treatment	(J) treatment	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
1	2	-761.200*	281.952	.017	-1365.93	-156.47
	3	-616.600*	265.827	.036	-1186.74	-46.46
	4	369.050	281.952	.212	-235.68	973.78
2	1	761.200*	281.952	.017	156.47	1365.93
	3	144.600	281.952	.616	-460.13	749.33
	4	1130.250*	297.203	.002	492.81	1767.69
3	1	616.600*	265.827	.036	46.46	1186.74
	2	-144.600	281.952	.616	-749.33	460.13
	4	985.650*	281.952	.004	380.92	1590.38
4	1	-369.050	281.952	.212	-973.78	235.68
	2	-1130.250*	297.203	.002	-1767.69	-492.81
	3	-985.650*	281.952	.004	-1590.38	-380.92

\*. The mean difference is significant at the 0.05 level.

The result is significant so we have to perform Post HOC test to determine the cause of difference. Here treatment 1 and 4 have similar effect and they are different from treatments 2 and 3.

## Using STATA

oneway value treatment

oneway value treatment

## Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	3561394.36	3	1187131.45	6.72	0.0049
Within groups	2473236.75	14	176659.768		
Total	6034631.11	17	354978.301		

Bartlett's test for equal variances: chi2(3) = 4.0595 Prob&gt;chi2 = 0.255

For pairwise comparison

oneway value treatment, bonferroni tabulate

oneway value treatment, bonferroni tabulate

Treatment	Summary of value		
	Mean	Std. Dev.	Freq.
1	1916.8	539.09016	5
2	2678	488.86194	4
3	2533.4	172.59287	5
4	1547.75	397.76574	4
Total	2175.2222	595.80055	18

Source	Analysis of Variance				
	SS	df	MS	F	Prob > F
Between groups	3561394.36	3	1187131.45	6.72	0.0049
Within groups	2473236.75	14	176659.768		
Total	6034631.11	17	354978.301		

Bartlett's test for equal variances:  $\chi^2(3) = 4.0595$  Prob> $\chi^2 = 0.255$

Comparison of value by treatment  
(Bonferroni)

Row Mean - Col Mean	1	2	3
2	761.2 0.104		
3	616.6 0.216	-144.6 1.000	
4	-369.05 1.000	-1130.25 0.012	-985.65 0.021

Two way ANOVA

The following table gives the result of the experiment on four varieties of a crop in 5 blocks of plot.

Block I	Block II	Block III	Block IV	Block V
A 32	B 33	D 30	A 35	C 36
B 34	C 34	C 35	C 32	D 29
C 31	A 34	B 36	B 37	A 37
D 29	D 26	A 33	D 28	B 35

Analyse the above result to test whether there is significant difference between yields of four varieties.

### Using EXCEL

Data\ Data analysis\ Anova: Two Factor without Replication

A	B	C	D	E	F	G
1 Treatment	Block 1	Block 2	Block 3	Block 4	Block 5	
2 A	32	34	33	35	37	
3 B	34	33	36	37	35	
4 C	31	34	35	32	36	
5 D	29	26	30	28	29	

A	B	C	D	E	F	G
8 Anova: Two-Factor Without Replication						
9						
10 SUMMARY	Count	Sum	Average	Variance		
11 A	5	171	34.2	3.7		
12 B	5	175	35	2.5		
13 C	5	168	33.6	4.3		
14 D	5	142	28.4	2.3		
15						
16 Block 1	4	126	31.5	4.333333		
17 Block 2	4	127	31.75	14.91667		
18 Block 3	4	134	33.5	7		
19 Block 4	4	132	33	15.33333		
20 Block 5	4	137	34.25	12.91667		
21						
22						
23 ANOVA						
24 Source of Variation						
25 Rows	SS	df	MS	F	P-value	F crit
26 Columns	134	3	44.66667	18.16949	9.3E-05	3.490295
27 Error	21.7	4	5.425	2.20678	0.129622	3.259167
28	29.5	12	2.458333			
29 Total	185.2	19				

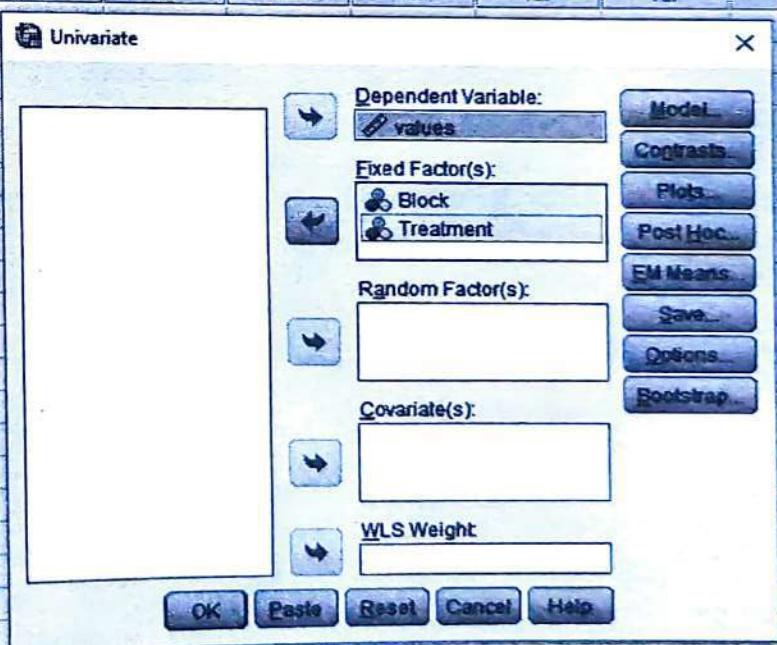
A	B	C	D	E	F	G	H
Total	185.2	19					
$\alpha$	0.05						
Decision				Formula			
the null hypothesis $H_0$ is rejected				=IF(E25<G25,"there is no reason to reject Null hypothesis $H_0$ ","the null hypothesis $H_0$ is rejected")			
there is no reason to reject Null hypothesis $H_0$				=IF(E26<G26,"there is no reason to reject Null hypothesis $H_0$ ","the null hypothesis $H_0$ is rejected")			
it is significant				=IF(F25<B31,"it is significant","it is not significant")			
it is significant				=IF(F26<B31,"it is significant","it is not significant")			

## Using SPSS

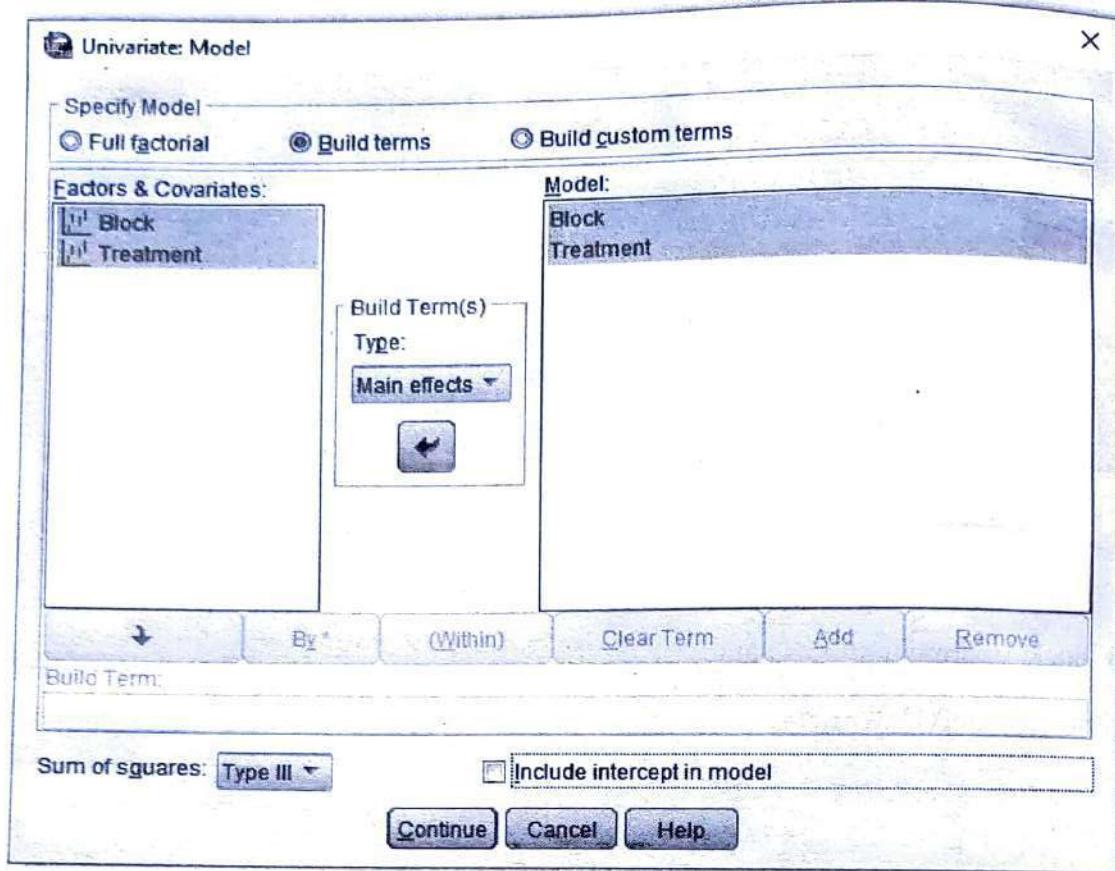
Create variable for block, treatment and its values

Analyze\General linear model\Univariate

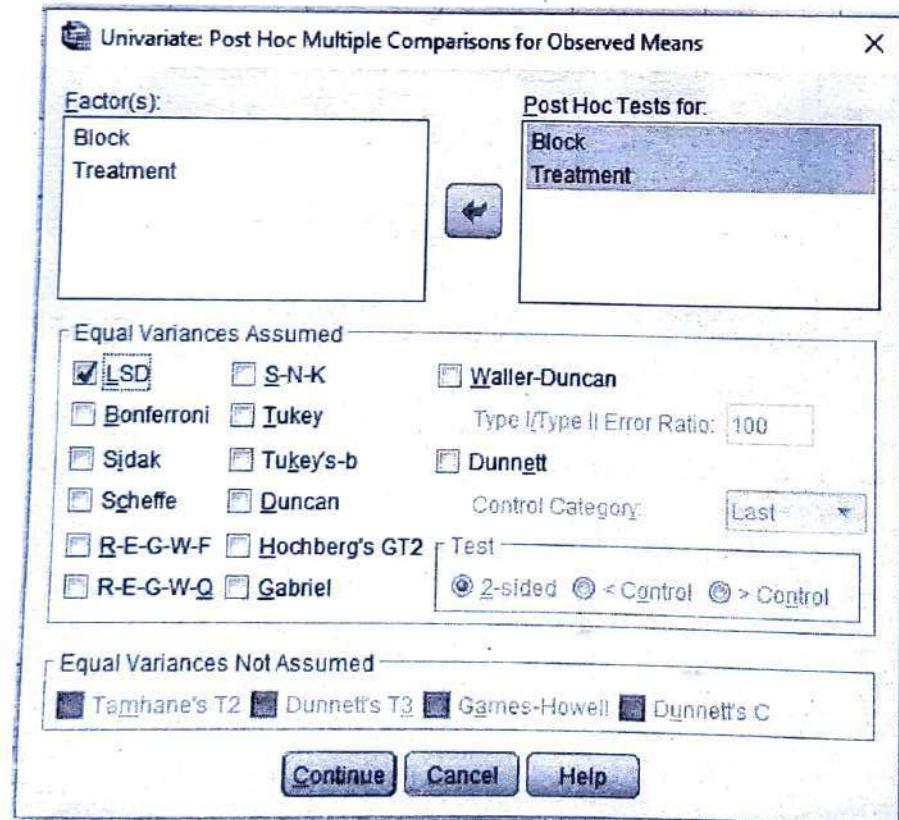
	Block	Treatment	values
1	1	A	32
2	1	B	34
3	1	C	31
4	1	D	29
5	2	A	34
6	2	B	33
7	2	C	34
8	2	D	26
9	3	A	33
10	3	B	36
11	3	C	35
12	3	D	30
13	4	A	35
14	4	B	37
15	4	C	32
16	4	D	28
17	5	A	37
18	5	B	35
19	5	C	36
20	5	D	29



## Click model



## Click Post HOC



outputs  
Block Block

Treatment Treatment

**Between-Subjects Factors**

		Value Label	N
	1		4
	2		4
	3		4
	4		4
	5		4
	1	A	5
	2	B	5
	3	C	5
	4	D	5

**Tests of Between-Subjects Effects**

Dependent Variable: values values

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	Model				
Model	21672.500 <sup>a</sup>	8	2709.062	1101.992	.000
Block	21.700	4	5.425	2.207	.130
Treatment	134.000	3	44.667	18.169	.000
Error	29.500	12	2.458		
Total	21702.000	20			

a. R Squared = .999 (Adjusted R Squared = .998)

**Multiple Comparisons**

Dependent Variable: values values

LSD

(I) Block	(J) Block	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.25	1.109	.825	-2.67	2.17
	3	-2.00	1.109	.096	-4.42	.42
	4	-1.50	1.109	.201	-3.92	.92
	5	-2.75*	1.109	.029	-5.17	-.33
	1	.25	1.109	.825	-2.17	2.67
2	3	-1.75	1.109	.140	-4.17	.67
	4	-1.25	1.109	.282	-3.67	1.17
	5	-2.50*	1.109	.044	-4.92	-.08
	1	2.00	1.109	.096	-.42	4.42
	2	1.75	1.109	.140	-.67	4.17
3	4	.50	1.109	.660	-1.92	2.92
	5	-.75	1.109	.512	-3.17	1.67
	1	1.50	1.109	.201	-.92	3.92
	2	1.25	1.109	.282	-1.17	3.67
	3	-.50	1.109	.660	-2.92	1.92
4	5	-1.25	1.109	.282	-3.67	1.17
	1	2.75*	1.109	.029	.33	5.17
	2	2.50*	1.109	.044	.08	4.92
	3	.75	1.109	.512	-1.67	3.17
	4	1.25	1.109	.282	-1.17	3.67

Based on observed means.

The error term is Mean Square(Error) = 2.458.

\* The mean difference is significant at the 0.05 level.

## Using STATA

anova values Block Treatment

anova values Block Treatment

Number of obs =	20	R-squared =	0.8407		
Root MSE =	1.56791	Adj R-squared =	0.7478		
Source	Partial SS	df	MS	F	Prob>F
Model	155.7	7	22.242857	9.05	0.0006
Block Treatment	21.7 134	4 3	5.425 44.666667	2.21 18.17	0.1296 0.0001
Residual	29.5	12	2.4583333		
Total	185.2	19	9.7473684		

## LSD

The following is the  $5 \times 5$  Latin square design for data taken from a manurial experiment with sugarcane. The five treatments were A=no manure; B= an inorganic manure; C,D and E =three levels of farm yard manure.

Plan and yield of sugarcane (in a suitable unit) per plot.

Row	Column				
	I	II	III	IV	V
I	A 52.5	E 46.3	D 44.1	C 48.1	B 40.9
II	D 44.2	B 42.9	A 51.3	E 49.3	C 32.6
III	B 49.1	A 47.3	C 38.1	D 41.0	E 47.2
IV	C 43.2	D 42.5	E 67.2	B 55.1	A 45.3
V	E 47.0	C 43.2	B 46.7	A 46.0	D 43.2

Analyse the above data to find if there are any treatment effects.

Using Excel

A	B	C	D	E	F	G	H	I
<b>Table to calculate sums</b>								
Row		1	2	3	4	5	Grand Total	formula
1	52.5	46.3	44.1	48.1	40.9	231.9	=SUM(B14:F14)	
2	44.2	42.9	51.3	49.3	32.6	220.3	=SUM(B15:F15)	
3	49.1	47.3	38.1	41	47.2	222.7	=SUM(B16:F16)	
4	43.2	42.5	67.2	55.1	45.3	253.3	=SUM(B17:F17)	
5	47	43.2	46.7	46	43.2	226.1	=SUM(B18:F18)	
Grand Total	236	222.2	247.4	239.5	209.2	1154.3		
formula	=SUM(B14:B18)	=SUM(C14:C18)	=SUM(D14:D18)	=SUM(E14:E18)	=SUM(F14:F18)	=SUM(G14:G18)		
<b>treatment sums</b>								
Treatments	Sums	formula						
1	242.4	=SUMIF(Table1[Treatment],A24,Table1[Values])						
2	234.7	=SUMIF(Table1[Treatment],A25,Table1[Values])						
3	205.2	=SUMIF(Table1[Treatment],A26,Table1[Values])						
4	215	=SUMIF(Table1[Treatment],A27,Table1[Values])						
5	257	=SUMIF(Table1[Treatment],A28,Table1[Values])						
Grand Total	1154.3	=SUM(B24:B28)						

A	B	C	D	E	F	G
	symbol	value	formula			
no. of rows	m	5	=COUNT(A4:A18)			
no. of columns	m	5	=COUNT(B13:F13)			
no. of treatments	m	5	=COUNT(A24:A28)			
Grand total	G	1154.3	=B29			
correction factor	C.F.	53296.3	=C35^2/C32^2			
total sum of square	TSS	977.17	=SUMSQ(B4:F18)-C36			
sum of square due to row	SSR	141.078	=SUMSQ(C14:G18)/C32)-C36			
sum of square due to column	SSC	183.758	=SUMSQ(B19:F19)/C33)-C36			
sum of square due to treatment	SST	348.238	=SUMSQ(B24:B28)/C34)-C36			
sum of square due to error	SSE	304.095	=C37-C38-C39-C40			
level of significance	$\alpha$	0.05				

	degrees of freedom	Sum of squares	Mean sum of square	F ratio	F tabulated	sig
source of variation					=F.INV.RT(	=F.DIST.RT(
				=D47/	\$C\$42,B47,	E47,B47,\$B\$
				\$D\$50	\$B\$50)	50)
	=C32-1	=C38	=C47/B47			
Row		4	141.078	35.2696	1.392	3.2591667
Column		4	183.758	45.9396	1.813	3.2591667
Treatment		4	348.238	87.0596	3.435	3.2591667
Error		12	304.095	25.341267		
Total		24	977.17			

	A	B	C	D	E	F	G	H	I	J
53	Decision									
54	Significant approach									
55										
56										
57										
58										
59	P value approach									
60										
61										
62										

- 1 There is no reason to reject Null Hypothesis of no significant difference between Rows, Hence it is accepted  
 2 There is no reason to reject Null Hypothesis of no significant difference between Rows, Hence it is accepted  
 3 Null hypothesis is rejected

- 1 It is insignificant  
 2 It is insignificant  
 3 It is significant

## Using SPSS

Create variable for rows, columns, treatment and its values

Analyze\General linear model\Univariate