

Analysis of Variance (ANOVA)

The ANOVA is a powerful statistical tool for tests of significance. The test of significance based on t-distribution is an adequate procedure only for testing the significance of the difference between two sample means. In a situation when we have two or more samples to consider at a time, an alternative procedure is needed for testing the hypothesis that all the samples have been drawn from the same population. For example, if three fertilizers are to be compared to find their efficacy, this could be done by a field experiment, in which each fertilizer is applied to 10 plots and then the 30 plots are later harvested with the crop yield being calculated for each plot. Now we have 3 groups of ten figures and we wish to know if there are any differences between these groups. The answer to this problem is provided by the technique of ANOVA.

The term ANOVA was introduced by Prof. R.A. Fisher in 1920's to deal with problem in the analysis of agronomical data. Variation is inherent in nature

The total variation in any set of numerical data is due to a number of causes which may be classified as i) Assignable causes and ii) chance causes

The variation due to assignable causes can be detected and measured where as the variation due to chance causes is beyond the control of humans and cannot be traced separately

ANOVA: The ANOVA is a simple arithmetical process of sorting out the components of variation in a given data

Types of ANOVA: There are two types i) One way classification and ii) Two way classification

Assumptions of ANOVA:

1. The observations are independent
2. Parent population from which observations are taken is normal
3. Various treatment and environmental effects are additive in nature
4. The experimental errors are distributed normally with mean zero and variance σ^2

Experimental Designs

In order to verify a hypothesis pertaining to some scientific phenomena we have to collect data. Such data are obtained by either observation or by experimentation. The main topics connected with data collection are Theory of Sample Surveys and Experimental Designs. In sample survey, a researcher makes observations on existing population and records data without interfering with the process that is being observed. In

experimentation, on the other hand, the researcher controls or manipulates the environment of the subjects that constitute the population. The experiments allow a researcher to study the factors of his interest and show that these factors actually cause certain effects. Hence, whenever the objective is to study the effects of variables rather than simply to describe a population, we prefer the data collection through experimentation.

Modern concepts of experimental design are due primarily to R.A. Fisher. He developed them in the planning of agricultural field experiments. They are now used in many fields of science.

Basic concepts:

Blocks: In agricultural experiments, most of the times we divide the whole experimental unit (field) into relatively homogeneous sub-groups or strata. These strata, which are more uniform amongst themselves than the field as a whole are known as blocks.

Treatments: the objects of comparison in an experiment are defined as treatments

For example: i) suppose an Agronomist wishes to know the effect of different spacings on the yield of a crop, different spacings will be treatments. Each spacing will be called a treatment.

ii) If different of fertilizer are tried in an experiment to test the responses of a crop to the fertilizer doses, the different doses will be treatments and each dose will be a treatment.

ii) A teacher practices different teaching methods on different groups in his class to see which yields the best results.

iii) A doctor treats a patient with a skin condition with different creams to see which is most effective.

Experimental unit: Experimental unit is the object to which treatment is applied to record the observations.

For example i) In laboratory insects may be kept in groups of five or six. To each group, different insecticides will be applied to know the efficacy of the insecticides. In this study different groups of insects will be the experimental unit.

ii) If treatments are different varieties, then the objects to which treatments are applied to make observations will be different plot of land. The plots will be called experimental units.

Basic principles of experimental designs:

The purpose of designing an experiment is to increase the precision of the experiment. In order to increase the precision, we try to reduce the experimental error. For reducing the experimental error, we adopt some techniques. These techniques form the basic principles of experimental designs. The basic principles of the experimental designs are replication, randomization and local control.

1. Replication: Repetition of treatment to different experimental units is known as Replication. In other words, the repetition of treatments under investigation is known as replication. We have no means of knowing about the variations in the results of a treatment. Only when we repeat the treatment several times we can estimate the experimental error.

A replication is used (i) to secure more accurate estimate of the experimental error, a term which represents the differences that would be observed if the same treatments were applied several times to the same experimental units;

(ii) to reduce the experimental error and thereby to increase precision, which is a measure of the variability of the experimental error.

The standard error of treatment mean is $\frac{\sigma}{\sqrt{r}}$. Where σ is S.D. of treatment in the population and 'r' is the number of replications. As 'r' increases, the standard error of mean decreases. Also in the analysis of variance the replication of treatments provides estimate of experimental error which is essential for the application of F-test.

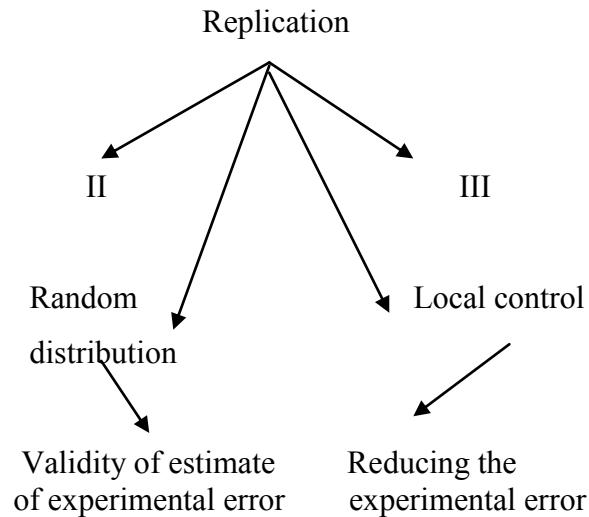
2. Randomization: when all the treatments have equal chances of being allocated to different experimental units it is known as randomization

or

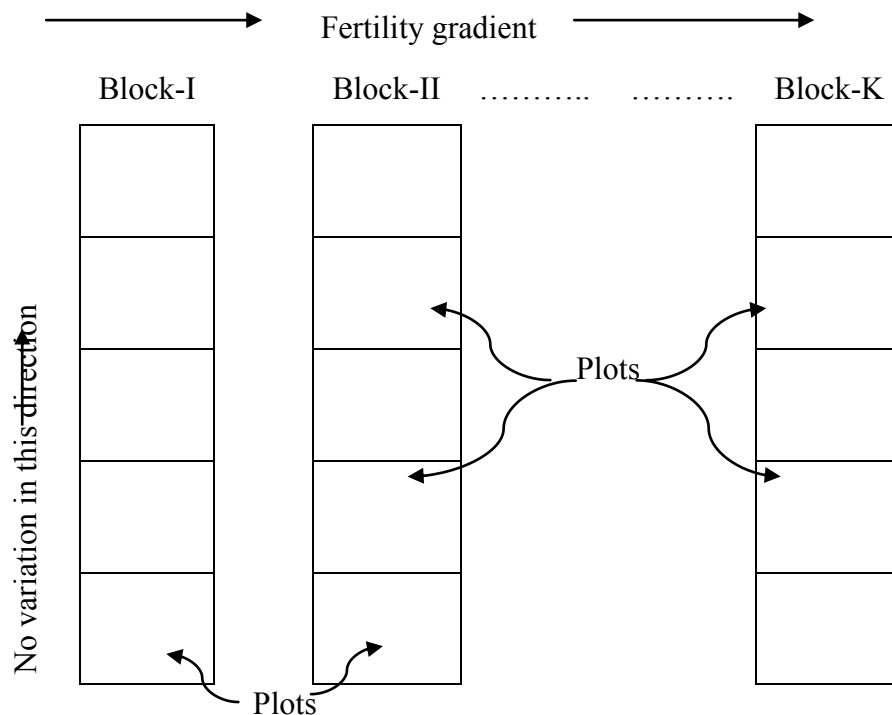
Random allocation of treatments to different experimental units known as randomization. The purpose of randomization is to remove bias and other sources of extraneous variation which are not controllable. Another advantage of randomization (accompanied by replication) is that it forms the basis of any valid statistical test. Hence the treatments must be assigned at random to the experimental units. Randomization is usually done by using tables of random numbers.

3. Local control: It has been observed that all extraneous sources of variation are not removed by randomization and replication. This necessitates a refinement in the experimental technique. For this purpose, we make use of local control, a term referring to the grouping of homogeneous experimental units.

The main purpose of the principle of local control is to increase the efficiency of an experimental design by decreasing the experimental error.



Shape of blocks and plots: the shape and size of the blocks will usually depend up on the shape and size of the plots. In order to control the experimental error it is desirable to divide the whole experimental area into different subgroups (blocks) such that within each block there is as much homogeneity as possible but between blocks there is maximum variation. Further each block is to be divided into as many plots as the number of treatments. For maximum precision the plots should be rectangular in shape with their long sides parallel to the direction of the fertility gradient and the blocks should be arranged one after the other along the fertility gradient as shown in the figure.



COMPLETELY RANDOMIZED DESIGN (CRD)

The CRD is the simplest of all the designs. In this design, treatments are allocated at random to the experimental units over the entire experimental material. In case of field experiments, the whole field is divided into a required number of plots equal size and then the treatments are randomized in these plots. Thus the randomization gives every experimental unit an equal probability of receiving the treatment.

In field experiments there is generally large variation among experimental plots due to soil heterogeneity. Hence, CRD is not preferred in field experiments. In laboratory experiments and green house studies, it is easy to achieve homogeneity of experimental materials. Therefore, CRD is most useful in such experiments.

Layout of CRD: The placement of the treatments on the experimental units along with the arrangement of experimental units is known as the layout of an experiment.

For example, suppose that there are 5 treatments A,B,C,D and E. each with 4 replications, we need 20 experimental units. Here, since the number of units is 20, a two digit random number of table will be consulted and a series of 20 random numbers will be taken excluding those which are greater than 20. suppose, the random numbers are 4,18,2,14,3,7,13,1,6,10,17,20,8,15,11,5,9,12,16,19. After this the plots will be serially numbered and the treatment A will be allotted to the plots bearing the serial numbers 4, 18, 2, 14 and so on.

1	2	3	4	5
B	A	B	A	D
6	7	8	9	10
C	B	D	E	C
11	12	13	14	15
D	E	B	A	D
16	17	18	19	20
E	C	A	E	C

Statistical analysis:

Let us suppose that there are 'k' treatments applied to 'r' plots. These can be represented by the symbols as follows:

Treatments	1.....2..... j.....n	Totals	means
t ₁	y ₁₁ y ₁₂ , ... y _{1j} ... y _{1r}	T ₁	\bar{T}_1
t ₂	y ₂₁ y ₂₂ , ... y _{2j} ... y _{2r}	T ₂	\bar{T}_2
.	.	.	.
.	.	.	.
t _i	y _{i1} y _{i2} , ... y _{ij} y _{ir}	T _i	\bar{T}_i
.	.	.	.
.	.	.	.
t _k	y _{k1} y _{k2} , ... y _{kj} ... y _{kr}	T _k	\bar{T}_k
		G.T.	

Mathematical model:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (i = 1, 2, \dots, k; \quad j = 1, 2, \dots, r)$$

Where y_{ij} is the j^{th} replication of the i^{th} treatment

μ = general mean effect

α_i = the effect due to i^{th} treatment = $[\bar{T}_i - \mu]$

ε_{ij} = error effect ($\varepsilon_{ij} \sim N(0, \sigma^2)$)

Null hypothesis H_0 : There is no significant difference between the treatment effects

$$\Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

Where $\alpha_i = \mu_i - \mu$, ($i = 1, 2, \dots, k$)

The null hypothesis can be verified by applying the ANOVA procedure. The steps involved in this procedure are as follows:

1) Correction factor = $\frac{(G.T.)^2}{N}$

2) Treatment Sum of Square (Tr.S.S.) = $\frac{(T_1^2 + T_2^2 + \dots + T_k^2)}{n} - CF$

$$= \frac{\sum_{i=1}^k T_i^2}{n} - CF$$

3) Total Sum of Square (TSS) = $\{y_{11}^2 + y_{12}^2 + y_{13}^2 + \dots + y_{kn}^2\} - CF$
 $= \sum \sum y_{ij}^2 - CF$

4) Error Sum of Square (ESS) = TSS - Tr.S.S.

ANOVA table

<i>Sources of variation</i>	<i>D.F.</i>	<i>S.S.</i>	<i>M.S.</i>	<i>F-cal.value</i>	<i>F-table value</i>
<i>Treatments</i>	$k-1$	<i>Tr.S.S.</i>	$TMS = \frac{Tr.S.S.}{k-1}$	$F_t = \frac{TMS}{EMS}$	$F[k-1, N-k]$ at $\alpha\%$ LOS
<i>Error</i>	$N-k$	<i>ESS</i>	$EMS = \frac{ESS}{N-k}$	-	
<i>Total</i>	$N-1$	<i>TSS</i>			

If the calculated value of $F <$ table value of F at certain level of significance, we accept H_0 and hence we may conclude that there is no significant difference between the treatment means

If the calculated value of $F >$ table value of F , H_0 is rejected. Then the problem is to know which of the treatment means are significantly different. For this, we calculate critical difference (CD)

$$CD = SED \times t - \text{table value for error d.f. at } 5\% \text{ LOS.}$$

where SED = Standard Error of Difference between the Treatments.

$$SED = \sqrt{\frac{2EMS}{r}} \quad (\text{equal No. of replications}), \text{ where } r \text{ is number of replications}$$

$$SED = \sqrt{EMS \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}, \text{ where } i = 1, 2, \dots, k \text{ and } j = 1, 2, \dots, k \text{ (unequal No. replications)}$$

The treatment means are arranged first in descending order of magnitude. If the difference between the two treatment means is less than CD value, it will be declared as non significant otherwise significant

Advantages and disadvantages of CRD:

1. This design is most commonly used in laboratory experiments such as in Ag. Chemistry, plant pathology, and animal experiments where the experimental material is expected to be homogeneous.
2. This design is useful in pot cultural experiments where the same type of soil is usually used. However, in greenhouse experiments care has to be taken with regard to sunshade, accessibility of air along and across the bench before conducting the experiment.

3. Any number of replications and treatments can be used. The number of replications may vary from treatment to treatment.
4. The analysis remains simple even if information on some units are missing
5. This design provides maximum number of degrees of freedom for the estimation of error than the other designs
6. The only draw back with this design is that when the experimental material is heterogeneous, the experimental error would be inflated and consequently the treatments are less precisely compared. The only way to keep the experimental error under control is to increase the number of replications thereby increasing the degrees of freedom for error.

Applications: 1. CRD is most useful in laboratory technique and methodological studies.

Ex: in physics, chemistry, in chemical and biological experiments, in some greenhouse studies etc.

2. CRD is also recommended in situations where an appreciable fraction of units is likely to be destroyed or fail to respond.

Case i) CRD with equal repetitions:

Example: In order to find out the yielding abilities of five varieties of sesamum an experiment was conducted in the greenhouse using a completely randomized design with four pots per variety. Analyze the data and state your conclusions

Seed yield of sesamum, g/pot

Varieties				
1	2	3	4	5
8	10	18	12	8
8	12	17	10	11
6	13	13	15	9
10	9	16	11	8

Sol: H_0 : There is no significant difference between effect varieties.

$$\text{i.e. } \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$$

$$\text{Correction factor (CF)} = \frac{(GT)^2}{N}$$

$$\text{Total sum of squares (TSS)} = \{y_{11}^2 + y_{12}^2 + y_{13}^2 + \dots + y_{45}^2\} - CF$$

Variety Sum of squares (VSS) =

$$\frac{(\sum v_1)^2 + (\sum v_2)^2 + (\sum v_3)^2 + (\sum v_4)^2 + (\sum v_5)^2}{r} - CF$$

Error Sum of Square (ESS) = TSS- VSS

		Varieties					
		1	2	3	4	5	
		8	10	18	12	8	
		8	12	17	10	11	
		6	13	13	15	9	
		10	9	16	11	8	
Variety totals		32	44	64	48	36	GT = 224
Means		8	11	16	12	9	

$$CF = \frac{(224)^2}{20} = 2508.8$$

$$TSS = \{(8)^2 + (10)^2 + (18)^2 + \dots + (16)^2 + (11)^2 + (8)^2\} - 2508.8 = 207.2$$

$$VSS = \frac{(32)^2 + (44)^2 + (64)^2 + (48)^2 + (36)^2}{4} - 2508.8 = 155.2$$

$$ESS = TSS - VSS = 207.2 - 155.2 = 52.0$$

ANOVA TABLE

Sources of variation	d.f.	S.S.	M.S.	F-cal value	F- table value
Varieties	5-1 =4	155.20	38.80	11.19	$F_{0.05}(4,15) = 3.06$
Error	19-4=15	52.0	3.47		
Total	20-1=19	207.20			

Calculated value of F > Table value of F at 5%LOS, H_0 is rejected and hence we conclude that there is significant difference between variety means.

Critical difference (CD): $SED \times t_{0.05}$ for error d.f.

$$\text{Where } SED = \sqrt{\frac{2EMS}{r}}$$

t-table value for 15 d.f at 5%LOS is 2.13

$$CD = \sqrt{\frac{2 * 3.47}{4}} \times 2.13 = 2.80$$

The varieties means are arranging descending order of magnitude. If the difference between the varieties mean is less than CD value, it will be declared as non significant otherwise significant

V ₃	V ₄	V ₂	V ₅	V ₁
16	12	11	9	8

The varieties which do not differ significantly have been underlined by a bar.

$$\text{Coefficient of variation (CV)} = \text{Coefficient of variation} = \frac{\sqrt{EMS}}{\bar{X}} \times 100$$

Where \bar{X} = Grand mean

$$\text{Coefficient of variation} = \frac{\sqrt{3.47}}{11.2} \times 100 = 16.6\%$$

Conclusion: Lesser CV% indicates more consistency in the data

Case –ii) CRD with unequal repetitions:

Example: A Completely Randomized Design was conducted with the three treatments A B and C where treatment A is replicated 6 times and B is replicated 4 times and C is replicated 5 times. Analyze the data and state your conclusions.

A	B	C
16.5	15.0	18.2
17.0	13.8	24.3
16.0	14.0	25.0
12.0	17.9	18.9
18.0	-	21.0
14.0	-	-

Sol: Null hypothesis H₀: There is no significant difference between the effect treatments

$$\text{i.e. } \alpha_1 = \alpha_2 = \alpha_3$$

First treatment A is replicated $r_1 = 6$ times

Second treatment B is replicated $r_2 = 4$ times

Third treatment C is replicated $r_3 = 5$ times

$$\text{Correction factor (CF)} = \frac{(GT)^2}{N}$$

$$\text{Total sum of squares (TSS)} = \{y_{11}^2 + y_{12}^2 + y_{13}^2 + \dots + y_{35}^2\} - CF$$

$$\text{Treatment of Sum of square (Tr.S.S.)} = \frac{(\sum A)^2}{r_1} + \frac{(\sum B)^2}{r_2} + \frac{(\sum c)^2}{r_3} - CF$$

$$\text{Error Sum of square(ESS)} = TSS - \text{Tr.S.S.}$$

$$\text{Standard error of difference} = \text{SED} = \sqrt{EMS \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$$

Where $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, r$

$$\text{Coefficient of variation} = \frac{\sqrt{2EMS}}{\bar{X}} \times 100; \quad \text{where } \bar{X} = \text{Grand mean}$$

Treatments							Totals	Means
A	16.5	17.0	16.0	12.0	18.0	14.0	$\Sigma A=93.5$	15.58
B	15.0	13.8	14.0	17.9	-	-	$\Sigma B=60.7$	15.18
C	18.2	24.3	25.0	18.9	21.0	-	$\Sigma C=107.4$	21.48
							GT=261.6	

$$CF = \frac{(261.6)^2}{15} = 4562.30$$

$$\begin{aligned} TSS &= (16.5)^2 + (17.0)^2 + (16.0)^2 + (12.0)^2 + (18.0)^2 + (14.0)^2 + (15.0)^2 + (13.8)^2 \\ &\quad + (14.0)^2 + (17.9)^2 + (18.2)^2 + (24.3)^2 + (25.0)^2 + (18.9)^2 + (21.0)^2 - 4562.30 \\ &= 4758.04 - 4562.30 \\ &= 195.74 \end{aligned}$$

$$\begin{aligned} Tr.S.S. &= \frac{(93.5)^2}{6} + \frac{(60.7)^2}{4} + \frac{(107.4)^2}{5} - 4562.30 \\ &= 4685.11 - 4562.30 \\ &= 122.81 \end{aligned}$$

$$\begin{aligned} ESS &= TSS - Tr.S.S. \\ &= 195.74 - 122.81 \\ &= 72.93 \end{aligned}$$

ANOVA TABLE

Sources	d.f	S.S.	M.S.	F-cal. Value	F- table Value
Treatments	3-1=2	122.81	61.405	10.10	$F_{0.05}(2,12) = 3.89$
Error	12	72.93	6.0775		
Total	15-1=14	195.74			

Calculated value $F(\text{Tr}) >$ Table value of F , H_0 is rejected and hence we conclude that there is significant difference between treatment means.

In CRD with unequal number of replications, we have to calculate CD for each pair of treatment means. As there are 3 treatments, we have to calculate $3C_2 = 3$ CD values.

Treatment pair	$SED = \sqrt{EMS \left(\frac{1}{r_i} + \frac{1}{r_j} \right)}$	CD = SED $\times t_{0.05}$ d.f.
AB	$\sqrt{6.08 \left(\frac{1}{6} + \frac{1}{4} \right)} = 1.59$	=1.59*2.18=3.47
AC	$\sqrt{6.08 \left(\frac{1}{6} + \frac{1}{5} \right)} = 1.49$	=1.49*2.18=3.25
BC	$\sqrt{6.08 \left(\frac{1}{4} + \frac{1}{5} \right)} = 1.65$	=1.65*2.18=3.61

Bar Notation:

The treatment means are arranged according to their ranks

T_3	T_1	T_2
21.48	15.58	15.18

- i) Those pairs not scored are significant
- ii) Those pairs under scored are non-significant

$$\text{Coefficient of variation} = \frac{\sqrt{EMS}}{\bar{X}} \times 100$$

where \bar{x} = Grand mean

$$= \frac{\sqrt{6.08}}{17.44} \times 100 = 14.14\%$$

Among three treatments, the third treatment i.e. T_3 is found to be superior one

RANDOMIZED BLOCK DESIGN (RBD)

We have seen that in a completely randomized design no local control measure was adopted excepting that the experimental units should be homogeneous. Usually, when experiments require a large number of experimental units, completely randomized designs cannot ensure precision of the estimates of treatment effects.

In agricultural field experiments, usually the experimental materials are not homogeneous. In such situations the principle of local control is adopted and the experimental material is grouped into homogeneous sub groups. The subgroup is

commonly termed as block. Since each block will consist the entire set of treatments a block is equivalent to a replication.

The blocks are formed with units having common characteristics which may influence the response under study. In agricultural field experiments the soil fertility is an important character that influences the crop responses. The uniformity trial is used to identify the soil fertility of a field. If the fertility gradient is found to run in one direction (say from north to south) then the blocks are formed in the opposite direction (from east to west).

If the number of experimental units within each group is same as the number of treatments and if every treatment appears precisely once in each group, then such an arrangement is called a randomized block design.

Layout: Let us consider 5 treatments A, B, C, D and E each replicated 4 times. We divide the whole experimental area into 4 relatively homogeneous blocks and each block into 5 plots. Treatments are then allocated at random to the plots of a block, fresh randomization being done for each block. A particular layout as follows.

Block-1	A	E	B	D	C
Block-2	E	D	C	B	A
Block-3	C	B	A	E	D
Block-4	A	D	E	C	B

Let us select one digit random numbers in the order of their occurrence in the table leaving 0 and greater than 5. suppose we get random numbers from 1 to 5 as: 1,3,5,4,2. So in the First block we allocate treatment A to the 1st plot and B to 3rd plot and so on.

Statistical Analysis: The results from RBD can be arranged in two way table according to the replications (blocks) and treatments; there will be 'rk' observations in total. The data can be arranged in the following table.

Treatments	Blocks b ₁b ₂b _jb _r	Treatment Totals	Means
t ₁	y ₁₁ y ₁₂ , ... y _{1j}y _{1r}	T ₁	\bar{T}_1
t ₂	y ₂₁ y ₂₂ , ... y _{2j}y _{2r}	T ₂	\bar{T}_2
⋮	⋮	⋮	⋮
t _i	y _{i1} y _{i2} , ... y _{ij} y _{ir}	T _i	\bar{T}_i
⋮	⋮	⋮	⋮
t _k	y _{k1} y _{k2} , ... y _{kj} y _{kr}	T _k	\bar{T}_k
Block totals	B ₁ B ₂B _jB _r	G.T.	
Means	\bar{B}_1 \bar{B}_2 \bar{B}_j \bar{B}_r		

Mathematical model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (i = 1, 2, \dots, k; j = 1, 2, \dots, r)$$

Where y_{ij} is the response of the j^{th} block and i^{th} treatment

μ = general mean effect

α_i = the effect due to i^{th} treatment

β_j = the effect due to j^{th} block

ε_{ij} is the error effect ($\varepsilon_{ij} \sim N(0, \sigma^2)$)

Null hypothesis: i) H_{01} : There is no significant difference between the treatment effects.

i.e. $\alpha_1 = \alpha_2 = \dots = \alpha_k$

ii) H_{02} : There is no significant difference between the block effects

i.e. $\beta_1 = \beta_2 = \dots = \beta_r$

The null hypothesis can be verified by applying the ANOVA procedure. The different steps are in the analysis of data are:

1) Correction factor = $\frac{(G.T)^2}{rk}$

2) Treatment Sum of Squares (Tr.S.S.) = $\frac{(T_1)^2 + (T_2)^2 + \dots + (T_k)^2}{r} - CF$

$$= \frac{\sum_{i=1}^k T_i^2}{r} - CF$$

$$3) \text{ Block Sum of Squares (BSS)} = \frac{(B_1^2) + (B_2^2) + \dots + (B_r^2)}{k} - CF$$

$$= \frac{\sum_{j=1}^r B_j^2}{k} - CF$$

$$4) \text{ Total Sum of Squares (TSS)} = \{y_{11}^2 + y_{12}^2 + y_{13}^2 + \dots + y_{kr}^2\} - CF$$

$$= \sum_{i=1}^k \sum_{j=1}^r y_{ij}^2 - CF$$

$$5) \text{ Error Sum of Square (ESS)} = TSS - Tr.S.S. - SSB$$

ANOVA TABLE

Sources of variation	D.F	S.S	M.S.	F-Cal. Value	F-table value At 5% LOS
Treatments	$k-1$	$Tr.S.S.$	$TMS = \frac{Tr.S.S.}{k-1}$	$F_t = \frac{TMS}{EMS}$	$F[k-1, \{(r-1)(k-1)\}]$
Blocks (Replications)	$r-1$	BSS	$BMS = \frac{BSS}{r-1}$	$F_b = \frac{BMS}{EMS}$	$F[r-1, \{(r-1)(k-1)\}]$
Error	$(r-1)(k-1)$	ESS	$EMS = \frac{ESS}{(r-1)(k-1)}$		
Total	$rk-1$	TSS			

If the calculated value of F (Treatments) < table value of F, we accept H_0 , and hence we may conclude that there is no significant difference between the treatment means.

If calculated value of F (Treatments) > table value of F, we reject H_0 and hence we may conclude that there is significant difference between the treatment means.

If the treatments are significantly different, the comparison of the treatments is carried out on the basis of Critical Difference (C.D.).

C.D. = $SED(Tr) \times t_{(r-1)(k-1)}$ at α level of significance

$$\text{Where } SED = \sqrt{\frac{2xEMS}{r}}$$

- i) F (Blocks) should be not significant, if the planning of experiment is well manner.
- ii) Desirable C.V. (%) in field experiment and lab experiment.

Advantages and disadvantages of RBD:

1. The principle advantage of RBD is that it increases the precision of the experiment. This is due to the reduction of experimental error by adoption of local control.
2. The amount of information obtained in RBD is more as compared to CRD. Hence, RBD is more efficient than CRD.
3. Flexibility is another advantage of RBD. Any number of replications can be included in RBD. If large number of homogeneous units are available, large number of treatments can be included in this design.
4. Since the layout of RBD involves equal replication of treatments, statistical analysis is simple. Even when some observations are missing of certain treatments, the data can be analysed by the use of missing plot technique.
5. When the number of treatments is increased, the block size will increase. If the block size is large it may be difficult to maintain homogeneity within blocks. Consequently, the experimental error will be increased. Hence, RBD may not be suitable for large number of treatments. But for this disadvantage, the RBD is a versatile design. It is the most frequently used design in agricultural experiments.
6. The optimum blocks size in field experiments is 21 plots. i.e. we can not compare treatments which are > 21 in RBD to preserve homogeneity of plots, within a block.

Example: The yields of 6 varieties of a crop in lbs., along with the plan of the experiment, are given below. The number of blocks is 5, plot of size is 1/20 acre and the varieties have been represented by A, B, C, D and E and analyze the data and state your conclusions

B-I	B 12	E 26	D 10	C 15	A 26	F 62
B-II	E 23	C 16	F 56	A 30	D 20	B 10
B-III	A 28	B 9	E 35	F 64	D 23	C 14
B-IV	F 75	D 20	E 30	C 14	B 7	A 23
B-V	D 17	F 70	A 20	C 12	B 9	E 28

Solution:

Null hypothesis H_{01} : There is no significant difference between variety means

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6$$

H_{02} : There is no significant difference between block means

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$

$$\text{Correction factor} = \frac{(GT)^2}{rk}$$

$$\text{Variety Sum of Squares due to varieties (VSS)} = \frac{\sum v_i^2}{r} - CF$$

$$\text{Block Sum of square (BSS)} = \frac{\sum b_j^2}{k} - CF$$

$$\text{Total sum of squares (TSS)} = \sum \sum y_{ij}^2 - CF$$

$$\text{Error Sum of Square (ESS)} = \text{TSS} - \text{VSS} - \text{BSS}$$

First rearrange the given data

Blocks	Varieties						Block totals	Means
	A	B	C	D	E	F		
B ₁	26	12	15	10	26	62	$\Sigma B_1 = 151$	25.17
B ₂	30	10	16	20	23	56	$\Sigma B_2 = 155$	25.83
B ₃	28	9	14	23	35	64	$\Sigma B_3 = 173$	28.83
B ₄	23	7	14	20	30	75	$\Sigma B_4 = 169$	28.17
B ₅	20	9	12	17	28	70	$\Sigma B_5 = 156$	26.00
Variety totals	$\Sigma A = 127$	$\Sigma B = 47$	$\Sigma C = 71$	$\Sigma D = 90$	$\Sigma E = 142$	$\Sigma F = 327$	$GT = 804$	-
Means	25.4	9.4	14.2	18	28.4	65.4	-	-

$$CF = \frac{(804)^2}{30} = 21547.2$$

$$\begin{aligned} VSS &= \frac{(127)^2 + (47)^2 + (71)^2 + (90)^2 + (142)^2 + (327)^2}{5} - 21547.2 \\ &= 31714.4 - 21547.2 = 10167.2 \end{aligned}$$

$$\begin{aligned} BSS &= \frac{(151)^2 + (155)^2 + (173)^2 + (169)^2 + (156)^2}{6} - 21547.2 \\ &= 21608.67 - 21547.2 = 61.47 \end{aligned}$$

$$TSS = (12)^2 + (26)^2 + (10)^2 + (15)^2 + \dots + (12)^2 + (9)^2 + (28)^2 - 21547.2$$

$$= 32194 - 21547.2$$

$$= 10646.8$$

$$\text{ESS} = \text{TSS} - \text{BSS} - \text{Tr.S.S.}$$

$$= 10646.8 - 61.47 - 10167.2$$

$$= 418.13$$

ANOVA TABLE

Sources of variation	d.f	S.S.	M.S.	F-cal. Value	F- table Value
Blocks	5-1=4	61.47	15.37	0.74	$F_{0.05}(4, 20) = 2.87$
Varieties	6-1=5	10167.2	2033.44	97.25	$F_{0.05}(5, 20) = 2.71$
Error	29-4-5= 20	418.13	20.91		
Total	30-1-29	10646.8			

Calculated value of F (Treatments) > Table value of F, H_0 is rejected and hence we conclude that there is highly significant difference between variety means.

$$\text{Where SEm} = \sqrt{\frac{EMS}{r}} = \sqrt{\frac{20.91}{5}} = 2.04$$

$$\text{SED} = \sqrt{2} * \text{SEm} = 1.414 * 2.04 = 2.88$$

Critical difference = SED x t-table value for error d.f. at 5% LOS

$$\begin{aligned} \therefore \text{CD} &= 2.88 * 2.09 \\ &= 6.04 \end{aligned}$$

$$\text{Coefficient of variation} = \frac{\sqrt{EMS}}{\bar{X}} \times 100 = \frac{\sqrt{20.91}}{26.8} \times 100 = 17\%$$

Bar Notation:

$$\begin{array}{cccccc} \bar{F} & \bar{E} & \bar{A} & \bar{D} & \bar{C} & \bar{B} \\ 65.4 & 28.4 & 25.4 & 18.0 & 14.2 & 9.40 \end{array}$$

- i) Those pairs not scored are significant
ii) Those pairs underscored are non-significant

Variety F gives significantly higher yield than all the other varieties; varieties D,C and B are on par and gives significantly higher yield than variety A.

LATIN SQUARE DESIGN (LSD)

When the experimental material is divided into rows and columns and the treatments are allocated such that each treatment occurs only once in a row and once in a column, the design is known as latin square design. In this design eliminating fertility variations consists in an experimental layout which will control variation in two perpendicular directions

[Latin square designs are normally used in experiments where it is required to remove the heterogeneity of experimental material in two directions. This design requires that the number of replications (rows) equal the number of treatments]. In LSD the number of rows and number of columns are equal. Hence the arrangement will form a square.

Layout of LSD: In this design the number of rows is equal to the number of columns and it is equal to the number of treatments. Thus in case of 'm' treatments, there have to be $m \times m = m^2$ experimental units (plots) arranged in a square so that each row as well as each column contain 'm' plots. The 'm' treatments are then allocated at random to these rows and columns in such a way that every treatment occurs once and only once in each row and each column such a layout is known as $m \times m$ L.S.D and is extensively used in agricultural experiments. The minimum and maximum number of treatments required for layout of LSD is 5 to 12.

In LSD the treatments are usually denoted by alphabets like A,B,C...etc. For a latin square with five treatments the arrangement may be as follows

Square –I

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

Square – II..... etc.,

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

Statistical analysis: the mathematical model for LSD is given by

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \quad (i = j = k = 1, 2, \dots, m)$$

where Y_{ijk} denote the response from the unit (plot) in the i^{th} row, j^{th} column and receiving the k^{th} treatment

μ = general mean effect

$\alpha_i = i^{\text{th}}$ row effect

$\beta_j = j^{\text{th}}$ column effect

$\gamma_k = k^{\text{th}}$ treatment effect;

$\epsilon_{ijk} =$ error component

we know that total variation = variation due to rows + variation due to columns +
Variation due to treatments + variation due to error

Null hypothesis (H_0) = There is no significant difference between Rows, Columns and Treatment effects.

i.e. i) $H_{01} : \alpha_1 = \alpha_2 = \dots = \alpha_m$

ii) $H_{02} : \beta_1 = \beta_2 = \dots = \beta_m$ and

iii) $H_{03} : \gamma_1 = \gamma_2 = \dots = \gamma_m$

The steps in the analysis of the data for verifying the null hypothesis are:

Different component variations can be calculated as follows:

$$1) C.F = \frac{(G.T)^2}{m^2}$$

$$2) \text{Row Sum of Squares (RSS)} = \frac{\sum_{i=1}^m r_i^2}{m} - C.F$$

$$3) \text{Column Sum of Squares (CSS)} = \frac{\sum_{j=1}^m c_j^2}{m} - C.F$$

$$4) \text{Treatment Sum of Squares (Tr..S.S).} = \frac{\sum_{k=1}^m t_k^2}{m} - C.F$$

$$5) \text{Total Sum of Squares (TSS)} = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m y_{ijk}^2 - C.F$$

$$6) \text{Error Sum of Squares (ESS)} = TSS - RSS - CSS - Tr.S.S.$$

ANOVA TABLE

Sources	D.F	S.S	M.S.	F-cal. value	F-table value at 5%LOS
Rows	$m-1$	RSS	$RMS = \frac{RSS}{m-1}$	$F_R = \frac{RMS}{EMS}$	$F_R[m-1, \{(m-1)(m-2)\}]$
Columns	$m-1$	CSS	$CMS = \frac{CSS}{m-1}$	$F_C = \frac{CMS}{EMS}$	"
Treatments	$m-1$	$Tr.S.S.$	$TMS = \frac{Tr.S.S.}{m-1}$	$F_T = \frac{TMS}{EMS}$	"
Error	$(m-1)(m-2)$	ESS	$EMS = \frac{ESS}{m-1}$		
Total	m^2-1	TSS	-		

If calculate value of $F(\text{Tr}) < \text{table value of } F \text{ at } 5\% \text{LOS}$, H_0 is accepted and hence we may conclude that there is no significance difference between treatment effects.

If calculate value of $F(\text{Tr}) > \text{table value of } F \text{ at } 5\% \text{LOS}$, H_0 is rejected and hence we may conclude that there is significance difference between treatments effects.

If the treatments are significantly different, the comparison of the treatments is carried out on the basis of Critical Difference (C.D.).

$$\text{C.D.} = \text{SED (Tr)} \times t_{(r-1)(k-1)} \text{ at } \alpha \text{ level of significance}$$

$$\text{Where SED} = \sqrt{\frac{2xEMS}{m}}, \text{ where } m = \text{number of rows}$$

If F is significant, the significance of any treatment contrast can be tested by using the CD value.

Advantages of Latin Square Design: 1) With two way grouping or stratification LSD controls more of the variation than C.R.D. or R.B.D.

2) L.S.D. is an incomplete 3-way layout. Its advantage over complete 3-way layout is that instead of m^3 experimental units only m^2 units are needed. Thus a 4x4 L.S.D. results in saving of $64-16 = 48$ observations over a complete 3-way layout.

3) The statistical analysis is simple though slightly complicated than for R.B.D. Even with missing data the analysis remains relatively simple.

4) More than one factor can be investigated simultaneously.

5) The missing observations can be analysed by using missing plot technique.

Example: An experiment on cotton was conducted to study the effect of foliar application of urea in combinations with insecticidal sprays in the cotton yield. Five treatments were tried in a 6x6 Latin Square Design. The layout plan and yield is given below:

T ₂ 4.9	T ₄ 6.4	T ₅ 3.3	T ₁ 9.5	T ₃ 11.8
T ₃ 9.3	T ₁ 4.0	T ₂ 6.2	T ₅ 5.1	T ₄ 5.4
T ₄ 7.0	T ₃ 15.4	T ₁ 6.5	T ₂ 6.0	T ₅ 4.6
T ₅ 5.3	T ₂ 7.6	T ₃ 13.2	T ₄ 8.6	T ₁ 4.9
T ₁ 9.3	T ₅ 6.3	T ₄ 11.8	T ₃ 15.9	T ₂ 7.6

Analyze the data and state your conclusions

Sol:

Null hypothesis: Rows, Columns and Treatments effects are equal

or

$$H_{01} : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 ;$$

$$H_{02} : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$

$$H_{03} : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$$

$$CF = \frac{(GT)^2}{m^2}$$

$$RSS = \frac{\sum_{i=1}^m R_i^2}{m} - CF$$

$$CSS = \frac{\sum_{j=1}^m C_j^2}{m} - CF$$

$$Tr.S.S = \frac{\sum_{k=1}^m T_k^2}{m} - CF$$

$$TSS = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m y_{ij}^2 - CF$$

$$ESS = TSS - RSS - CSS - Tr.S.S.$$

ANOVA Table

Rows	Columns					Row totals
	1	2	3	4	5	
1	T ₂ 4.9	T ₄ 6.4	T ₅ 3.3	T ₁ 9.5	T ₃ 11.8	R ₁ = 35.9
2	T ₃ 9.3	T ₁ 4.0	T ₂ 6.2	T ₅ 5.1	T ₄ 5.4	R ₂ = 30.0
3	T ₄ 7.0	T ₃ 15.4	T ₁ 6.5	T ₂ 6.0	T ₅ 4.6	R ₃ = 9.5
4	T ₅ 5.3	T ₂ 7.6	T ₃ 13.2	T ₄ 8.6	T ₁ 4.9	R ₄ = 39.6
5	T ₁ 9.3	T ₅ 6.3	T ₄ 11.8	T ₃ 15.9	T ₂ 7.6	R ₅ = 50.9
Column totals	C ₁ = 35.8	C ₂ = 39.7	C ₃ = 41.0	C ₄ = 45.1	C ₅ = 34.3	GT=195.9

Treatment Totals: $\Sigma T_1 = 34.2$; $\Sigma T_2 = 32.3$; $\Sigma T_3 = 65.6$; $\Sigma T_4 = 39.2$; $\Sigma T_5 = 24.6$

$$CF = \frac{(195.9)^2}{25} = 1535.07$$

$$RSS = \frac{R_1^2 + R_2^2 + \dots + R_5^2}{5} - CF$$

$$= \frac{(35.9)^2 + (30.0)^2 + (39.5)^2 + (39.6)^2 + (50.9)^2}{5} - 1535.07$$

$$= 46.54$$

$$CSS = \frac{C_1^2 + C_2^2 + \dots + C_5^2}{5} - CF$$

$$= \frac{(35.8)^2 + (39.7)^2 + (41.0)^2 + (45.1)^2 + (34.3)^2}{5} - 1535.07$$

$$= 14.77$$

$$Tr.S.S = \frac{(\Sigma T_1)^2 + (\Sigma T_2)^2 + \dots + (\Sigma T_5)^2}{5} - CF$$

$$= \frac{(34.2)^2 + (32.3)^2 + ((65.6)^2 + (39.2)^2 + (24.6)^2)}{5} - 1535.07$$

$$= 196.55$$

$$\begin{aligned} \text{TSS} &= Y_{11}^2 + Y_{12}^2 + Y_{13}^2 + \dots + Y_{55}^2 - CF \\ &= (4.9)^2 + (6.4)^2 + (3.3)^2 + \dots + (15.9)^2 + (7.6)^2 - 1535.07 \\ &= 1821.07 - 1535.07 = 286.0 \end{aligned}$$

$$\begin{aligned} \text{ESS} &= \text{TSS} - \text{RSS} - \text{CSS} - \text{Tr.S.S} \\ &= 286.0 - 46.54 - 14.77 - 196.55 = 28.14 \end{aligned}$$

ANVOA table

Sources	d.f.	S.S.	M.S.	F cal . value	F table value
Rows	5-1 = 4	46.54	11.64	4.95	$F_{0.05}(4, 12) = 3.26$
Columns	5-1 = 4	14.77	3.69	1.57	$F_{0.05}(4, 12) = 3.26$
Treatments	5-1 = 4	196.55	49.14	20.91	$F_{0.05}(4, 12) = 3.26$
Error	24 - 12 = 12	28.14	2.35		
Total	25-1 = 24	286.0			

Calculated value of F (treatments) > Table value of t at 5% LOS, H_0 is rejected and hence effect of foliar application of urea, have significant effect in the yield. To determine which of the treatment pairs differ significantly we have to calculate the critical difference (C.D.)

$$\text{SEm} = \sqrt{\frac{\text{EMS}}{m}} = \sqrt{\frac{2.35}{5}} = 0.69$$

$$\text{SED} = \sqrt{2} * \text{SEm} = 1.414 * 0.69 = 0.97$$

$$\text{CD} = \text{SED} \times t - \text{table value for 12 d.f. at 5 \% LOS}$$

$$= 0.98 * 2.18$$

$$= 2.11$$

$$\text{Coefficient of variation (CV)} = \frac{\sqrt{\text{EMS}}}{\bar{X}} * 100 = \frac{\sqrt{2.35}}{7.84} * 100 = 19.55\%$$

Bar Notation:

\bar{T}_3	\bar{T}_4	\bar{T}_1	\bar{T}_2	\bar{T}_5
13.12	7.84	6.84	6.46	4.92

-
- i) The pairs not scored are significant
 - ii) The pairs under scored are non significant

From the bar chart it can be concluded that third treatment i.e. T_3 significantly higher than all the other treatments.

References Books:

1. Statistics for Agricultural Sciences, G. Nageswara Rao, Second Edition, BS Publications, Hyderabad
2. A Text book of Agricultural Statistics, R. Rangaswamy, New Age International (P) Limited, publishers
3. Statistical Methods, K.P. Dhamu and K. Ramamoorthy, AGROBIOS (INDIA)
4. Fundamentals of Mathematical Statistics, S.C. Gupta and V.K. Kapoor, Sultan Chand & Sons Educational Publications
5. Fundamentals Applied Statistics, S.C. Gupta and V.K. Kapoor, Sultan Chand & Sons Educational Publications
6. Design Resources Server: www.iasri.res.in