Module-5 Design of experiments and Reliability

Design of experiments:

A statistical experiment in any field is conducted to verify the truth of a particular hypothesis. Design of experiments may be defined as "the logical construction of the experiment in which the degree uncertainty with which the inference is drawn may be well defined".

Suppose we conduct an agricultural experiment to verify the truth of the claim that a particular fertilizer increases the yield of wheat. Design of experiments is used to control the "*extraneous variables*" (indirect variables involved. Eg. Fertility of soil in different plots, quantity of seeds used) and to minimize the experimental error so that the results of our experiments could be attributed to the "*experimental variables*" (direct variables involved. Eg. Quantity of fertilizer used and quantity of the yield) only.

The purpose of experimental design is to obtain maximum information of an experiment with the minimum experimental error, cost and labour.

Terms involved in experimental design:

1. **Treatments:** Various objects of comparison in a comparative experiment are called treatments. For example, in an agricultural experiment, different fertilizers used, or different types of crops or different methods of cultivation are the treatments.
2. **Experimental Units:** The smallest division of the experimental material to which we apply the treatments and on which we make observations on the experimental variables under the investigation is called experimental unit. In an agricultural experiment, the plot of land is the experimental unit.
3. **Blocks:** The whole experimental units are subdivided into homogenous subgroups. Such subgroups which are more homogenous among themselves are called blocks.
4. **Experimental error:** The error due to "*extraneous variables"* is called Experimental error.

Basic principles of experimental design:

The following three basic principles are adopted when the experiments are designed.

1. Randomisation.
2. Replication.
3. Local control.

1. **Randomisation:** Though our desire is to eliminate completely the effect of "*extraneous variables"* in any experiment, it is not possible. So, we try to control them by Randomisation. For example, in an agricultural experiment, to verify the truth of the claim that a particular fertilizer increases the yield, the selection of plots for experimental groups(plots in which fertilizers are used) and control groups (plots in which fertilizers are not used) is done in a random manner. This will eliminate the unknown bias in the experiment.
2. **Replication:** Replication means repetition. In an agricultural experiment, in which the effects of different fertilizers on the yield are studied, each fertilizer is used in more than one experimental plot. To estimate the quantity of experimental error, it is necessary to use each fertilizer for more than one plot so that we can get the precision of the effects of fertilizer.
3. **Local control:** To control the effect of "*extraneous variables"* we use local control which involves "*grouping"*(combining homogenous set of plots) , "*blocking"*(assigning same number of plots in different blocks) and "*balancing"*(adjusting grouping and blocking procedure) of experimental units.

Basic Design of experiments:

1. Completely Randomised Design (C.R.D)
2. Randomised Block Design (R.B.D)
3. Latin Square Design (L.S.D)
4. 2^2-Factorial Design

Completely Randomised Design (C.R.D):

Suppose that there are *N* plots available & we wish to compare *h* treatments for an experiment and let the i^{th} treatment be repeated (replicated) n_i times such that $n_1 + n_2 + \ldots + n_h = N$. The *N* plots are selected randomly and each given a treatment to get the design which is completely random. This design is used only if plots used are homogenous.

ANALYSIS OF VARIANCE (ANOVA):

There are situations which leads us to tests of hypotheses that involve more than two populations and information's from more than two samples. ANALYSIS OF VARIANCE (ANOVA) is a technique or procedure to test for the significance of the difference between more than two samples. The aim of ANOVA is to find how much total variability is due to each factor and it enables us to test for the significance of the difference among more than two sample means.

Assumptions of ANOVA:

1. Each sample is a random sample.
2. Each sample is independent of the other.
3. The samples are drawn from normal population.
4. Variances of the population are equal.

Basic steps of ANOVA:

1. Determine one estimate of the population variance from the variance among the sample means.
2. Determine a second estimate of the population variance from the variance within the sample.
3. Compare these two estimates; if they are approximately equal in value then accept null hypothesis.

Analysis of variance (ANOVA) for one way classification (or) one factor experiments:

Procedure:

Let N-Total number of items in the given data.

K- Number of treatment.

n- Number of samples in each treatment.

STEP 1: Set up Null Hypothesis H_0.

Prepare Table – I using the following steps:

STEP 2: Find Sum of all the items (T) of the sample $T = \sum T_1 + \sum T_2 + \cdots + \sum T_K$

STEP 3: Find Sum of squares of all the items of each treatment $S.S = \sum_1 {X_{ij}}^2 + \sum_2 {X_{ij}}^2 + \cdots + \sum_k {X_{ij}}^2$

STEP 4: Find the Correction Factor (C.F) = $\frac{T^2}{N}$

Table – I (calculations for ANOVA):

Classes	Values – Number of Blocks				Row total T_i	$\frac{{T_i}^2}{n_i}$	S.S = $\sum\sum {X_{ij}}^2$
	1	2	…	n			
1	x_{11}	x_{12}	…	x_{1n}	$T_1 = \sum_1 x_{ij}$	$\frac{{T_1}^2}{n_1}$	$\sum_1 {X_{ij}}^2$
2	x_{21}	x_{22}	…	x_{2n}	$T_2 = \sum_2 x_{ij}$	$\frac{{T_2}^2}{n_2}$	$\sum_2 {X_{ij}}^2$
…			…		…		…
i	x_{i1}	x_{i2}	…	x_{in}	$T_i = \sum_i x_{ij}$	$\frac{{T_i}^2}{n_i}$	$\sum_i {X_{ij}}^2$
…			…		…		…
k	x_{k1}	x_{k2}	…	x_{kn}	$T_k = \sum_k x_{ij}$	$\frac{{T_k}^2}{n_k}$	$\sum_k {X_{ij}}^2$
	C.F = $\frac{T^2}{N}$				**Grand total** $T = \sum T_i$	$\sum \frac{{T_i}^2}{n_i}$	$\sum\sum {X_{ij}}^2$

STEP 5: Find Total sum of squares (T.S.S) = Sum of squares of all the items - Correction Factor

$$= S.S - \frac{T^2}{N}$$

STEP 6: Find Sum of squares between samples (SSB) = $\frac{T_1^{\,2}}{n_1} + \frac{T_1^{\,2}}{n_1} + \ldots + \frac{T_1^{\,2}}{n_1} - \text{C.F}$

STEP 7: Find Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f(\text{Degrees of freedom})}$;

STEP 8: Find Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)

STEP 9: Find Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSB)}}{D.f(\text{Degrees of freedom})}$;

<u>Prepare Table-II (ANOVA table) as follows:</u>

STEP 10: Table-II (ANOVA Table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Between samples	SSB	$v_1 = K - 1$	$\text{MSB} = \frac{\text{SSB}}{\text{K}-1}$	$\text{F}_\text{C} = \frac{\text{MSB}}{\text{MSW}}$
Within samples	SSW	$v_2 = N - K$	$\text{MSW} = \frac{\text{SSW}}{\text{N}-\text{K}}$	$\text{F}_\text{C} = \frac{\text{MSW}}{\text{MSB}}$

Where,

SSB = **Sum of squares between samples**

SSW = **Sum of squares within samples**

MSB = **Mean square between samples**

MSW= **Mean square within samples**

$\text{F}_\text{C} =$ **Calculated F – value**

$= \frac{\text{MSB}}{\text{MSW}}$ If $\text{MSB} > MSW$ (or)

$= \frac{\text{MSW}}{\text{MSB}}$ If $\text{MSW} > MSB$

STEP 11:

1. If calculated value F_C > tabulated value F_T at 5% l.o.s for (v_1, v_2) D.f then the null hypothesis H_0 is rejected.
2. If calculated value F_C < tabulated value F_T at 5% l.o.s for (v_1, v_2) D.f then the null hypothesis H_0 is accepted.

NOTE:

If the values are larger we can simplify the calculations by shifting the origin. Shifting the origin does not affect the variance of a set of values, so we can take the origin at $x = x_0$ (say).

Problems:

<u>Type-1: Reframe the table by applying the principles of experimental design.</u>

1. **A completely randomized design experiment with 10 plots and 3 experiments gave the following result.**

Plot No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Treatments	**A**	**B**	**C**	**A**	**C**	**C**	**A**	**B**	**A**	**B**
Yield	**5**	**4**	**3**	**7**	**5**	**1**	**3**	**4**	**1**	**7**

Analyse the result for treatment effects.

Solution:

Let us reframe the table by applying the principles of experimental design.

Treatments K=3	**Yield of the plots**				**n**
	Number of Blocks				
	1	**2**	**3**	**4**	
A	5	7	3	1	4
B	4	4	7	-	3
C	3	5	1	-	3

N-Total number of items in the given data = 10.

K- Number of Treatments = 3.

n- Number of samples in each treatment (not equal in this problem).

Null Hypothesis H_0: There is no significant difference among the average yields in the 3 treatments.

Table – I (calculations for ANOVA):

Treatments K=3	**Yield of the plots / Number of Blocks**				**Row total T_i**	$\frac{T_i^2}{n_i}$	$\sum X_{ij}^2$
	1	**2**	**3**	**4**			
A	5	7	3	1	16	$\frac{16^2}{4} = 64$	84
B	4	4	7	-	15	$\frac{15^2}{3} = 75$	81
C	3	5	1	-	9	$\frac{9^2}{3} = 27$	35
	C.F = $\frac{T^2}{N} = \frac{40^2}{10} = 160$				**Grand total** $T = \sum T_i = 40$	$\sum \frac{T_i^2}{n_i} = 166$	$\sum\sum X_{ij}^2 = 200$

Total sum of squares (T.S.S) = Sum of squares of all the items ($\sum\sum X_{ij}^2$) - Correction Factor(C.F)

$= 200 - 160 = 40$

Sum of squares between samples (SSB) = $\sum \frac{T_i^2}{n_i}$ - Correction Factor(C.F) = $166 - 160$ = 6

Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f \to \text{K}-1} = \frac{6}{3-1} = \frac{6}{2} = 3$

Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)

= 40 – 6 = 34

Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSW)}}{D.f \to \text{N}-\text{K}} = \frac{34}{10-3} = \frac{34}{7} = 4.86$

Table-II (ANOVA table):

Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F-Ratio**
Between samples	SSB = 6	$\upsilon_1 = 3 - 1 = 2$	$\text{MSB} = \frac{\text{SSB}}{\text{K}-1} = 3$	$\text{F}_\text{C} = \frac{\text{MSW}}{\text{MSB}} = 1.62$
Within samples	SSW = 34	$\upsilon_2 = 10 - 3 = 7$	$\text{MSW} = \frac{\text{SSW}}{\text{N}-\text{K}} = 4.86$	

∴ Calculated value of F is 1.62

Tabulated value of F for (7, 2) ($\text{MSW} > MSB$)at 5% l.o.s is 19.35.

∵ $\text{F}_\text{C} < \text{F}_\text{T}$ Null hypothesis H_0 is accepted.

2. **Three varieties A, B, C of a crop are tested in a completely randomized design with four replications, the layout being given in the following table. The plot yields in pounds are also indicated. Analyze the experimental yield and state your conclusion.**

A(6)	**C(5)**	**A(8)**	**B(9)**
C(8)	**A(4)**	**B(6)**	**C(9)**
B(7)	**B(6)**	**C(10)**	**A(6)**

Solution:

Let us reframe the table by applying the principles of experimental design.

Varieties K=3	**Yield of the plots / Number of Blocks**				**n**
	1	**2**	**3**	**4**	
A	6	8	4	6	4
B	9	6	7	6	4
C	5	8	9	10	4

N-Total number of items in the given data = 12.
K- Number of Varieties = 3.
n- Number of samples in each treatment = 4(equal).

Null Hypothesis H_0: There is no significant difference among the average yields in the 3 treatments.

Table – I (calculations for ANOVA):

Treatments K=3	Yield of the plots / Number of Blocks				Row total T_i	$\frac{T_i^2}{n_i}$	$\sum X_{ij}^2$
	1	2	3	4			
A	6	8	4	6	24	$\frac{24^2}{4} = 144$	152
B	9	6	7	6	28	$\frac{28^2}{4} = 196$	202
C	5	8	9	10	32	$\frac{32^2}{4} = 256$	270
	C.F = $\frac{T^2}{N} = \frac{84^2}{12} = 588$				**Grand total** $T = \sum T_i = 84$	$\sum \frac{T_i^2}{n_i} = 596$	$\sum\sum X_{ij}^2 = 624$

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 624 - 588 = 36$

Sum of squares between samples (SSB) = $\sum \frac{T_i^2}{n_i}$ - Correction Factor(C.F) = 596 − 588 = 8

Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f \to \text{K}-1} = \frac{8}{3-1} = \frac{8}{2} = 4$

Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)

= 36 – 8 = 28

Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSW)}}{D.f \to \text{N}-\text{K}} = \frac{28}{12-3} = \frac{28}{9} = 3.11$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Between samples	SSB = 8	$\upsilon_1 = 3 - 1 = 2$	$\text{MSB} = \frac{\text{SSB}}{\text{K}-1} = 4$	$\text{F}_\text{C} = \frac{\text{MSB}}{\text{MSW}} = 1.286$
Within samples	SSW = 28	$\upsilon_2 = 12 - 3 = 9$	$\text{MSW} = \frac{\text{SSW}}{\text{N}-\text{K}} = 3.11$	

∴ Calculated value of F is 1.286

Tabulated value of F for (2, 9) at 5% l.o.s is 4.26.

∵ $\text{F}_\text{C} < \text{F}_\text{T}$ Null hypothesis H_0 is accepted.

Type-2: No reframe required:

3. **In order to determine whether there is significant difference in the durability of 3 makes of computers, samples of size 5 are selected from each make and the frequency of repair during 1st year of purchase is observed. The results are as follows:**

Makes	**I**	**5**	**6**	**8**	**9**	**7**
	II	**8**	**10**	**11**	**12**	**4**
	III	**7**	**3**	**5**	**4**	**1**

In the view of above data what conclusion can you draw?

Solution:

N-Total number of items in the given data = 15.
K- Number of Makes = 3.
n- Number of samples in each make = 5.

Null Hypothesis H_0: There is no significant difference in the durability of 3 makes of computers.

Table – I (calculations for ANOVA):

Makes K=3	**Number of samples (n = 5)**					**Row total T_i**	$\frac{T_i^2}{n_i}$	$\sum X_{ij}^2$
	1	**2**	**3**	**4**	**5**			
A	5	6	8	9	7	35	$\frac{35^2}{5} = 245$	255
B	8	10	11	12	4	45	$\frac{45^2}{5} = 405$	445
C	7	3	5	4	1	20	$\frac{20^2}{5} = 80$	100
	C.F = $\frac{T^2}{N} = \frac{100^2}{15} = 666.67$					**Grand total** $T = \sum T_i = 100$	$\sum \frac{T_i^2}{n_i} = 730$	$\sum\sum X_{ij}^2 = 800$

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 800 - 666.67 = 133.33$

Sum of squares between samples (SSB) = $\sum \frac{T_i^2}{n_i}$ - Correction Factor(C.F) = $730 - 666.67$

= 63.33

Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f \to \text{K}-1} = \frac{63.33}{3-1} = \frac{63.33}{2}$

$= 31.67$

Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)

= 133.33 – 63.33= 70

Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSW)}}{D.f \to \text{N}-\text{K}} = \frac{70}{15-3} = \frac{70}{12}$

$= 5.83$

Table-II (ANOVA table):

Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F-Ratio**
Between samples	SSB = 63.33	$\upsilon_1 = 3 - 1 = 2$	$\text{MSB} = \frac{\text{SSB}}{\text{K}-1} = 31.67$	$\text{F}_\text{C} = \frac{\text{MSB}}{\text{MSW}} = 5.43$
Within samples	SSW = 70	$\upsilon_2 = 15 - 3 = 12$	$\text{MSW} = \frac{\text{SSW}}{\text{N}-\text{K}} = 5.83$	

∴ Calculated value of F is 5.43

Tabulated value of F for (2, 12) at 5% l.o.s is 3.88.

∵ $\text{F}_\text{C} > \text{F}_\text{T}$ Null hypothesis H_0 is rejected.

4. **To test the significance of the variation of the retail prices of a certain commodity in the four principal states A, B, C, D seven shops were chosen at random in each city and the prices observed as follows:**

A	**82**	**79**	**73**	**69**	**69**	**63**	**61**
B	**84**	**82**	**80**	**79**	**76**	**68**	**62**
C	**88**	**84**	**80**	**68**	**68**	**66**	**66**
D	**79**	**77**	**76**	**74**	**72**	**68**	**64**

Do the data indicate that the prices in the four cities are significantly different?

Solution:

N-Total number of items in the given data = 28.

K- Number of states = 4.

n- Number of shops in each state= 7.

Null Hypothesis H_0: There is no significant difference among the prices in the 4 states.
Shifting the origin does not affect the variance of a set of values, so we take the origin at (say) x = 76 (to make the calculations simple).

Table – I (calculations for ANOVA):

States K=4	Shops(n=7)							Row total T_i	$\frac{T_i^2}{n_i}$	$\sum X_{ij}^2$
	1	2	3	4	5	6	7			
A	6	3	-3	-7	-7	-13	-15	-36	$\frac{(-36)^2}{7} = 185.14$	546
B	8	6	4	3	0	-8	-14	-1	$\frac{(-1)^2}{7} = 0.14$	385
C	12	8	4	-8	-8	-10	-10	-12	$\frac{(-12)^2}{7} = 20.57$	552
D	3	1	0	-2	-4	-8	-12	-22	$\frac{(-22)^2}{7} = 69.14$	238
	C.F = $\frac{T^2}{N} = \frac{(-71)^2}{28} = 180.03$							Grand total $T = \sum T_i = -71$	$\sum \frac{T_i^2}{n_i} = 274.99$	$\sum\sum X_{ij}^2 = 1721$

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)
$= 1721 - 180.03 = 1540.97$

Sum of squares between samples (SSB) = $\sum \frac{T_i^2}{n_i}$ - Correction Factor(C.F) = $274.99 - 180.03$
= 94.96

Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f \to \text{K}-1} = \frac{94.96}{4-1} = \frac{94.96}{3}$
$= 31.65$

Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)
= 1540.97 – 94.96 = 1446.01

Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSW)}}{D.f \to \text{N}-\text{K}} = \frac{1446.01}{28-4} = \frac{1446.01}{24}$
$= 60.25$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Between samples	SSB = 94.96	$v_1 = 4 - 1 = 3$	$\text{MSB} = \frac{\text{SSB}}{\text{K}-1} = 31.65$	$\text{F}_\text{C} = \frac{\text{MSW}}{\text{MSB}} = 1.90$
Within samples	SSW = 1446.01	$v_2 = 28 - 4 = 24$	$\text{MSW} = \frac{\text{SSW}}{\text{N}-\text{K}} = 60.25$	

∴ Calculated value of F is 1.90
Tabulated value of F for (24, 3) $(\text{MSW} > MSB)$ at 5% l.o.s is 8.64.
∵ $\text{F}_\text{C} < \text{F}_\text{T}$ Null hypothesis H_0 is accepted.

5. **The following are the numbers of mistakes made in 5 successive days of 4 technicians working in photographic laboratory.**

Technicians I	Technicians II	Technicians III	Technicians IV
6	14	10	9
14	9	12	12
10	12	7	8
8	10	15	10
11	14	11	11

Test at 1% l.o.s whether the differences among the sample means can be attributed to chance.

Solution:

N-Total number of items in the given data = 20.

K- Number of Technicians = 4.

n- Number of errors of each Technician =5.

Null Hypothesis H_0: There is no significant difference in the mean errors committed by the 4 Technicians.

Shifting the origin does not affect the variance of a set of values, so we take the origin at (say) x = 10 (to make the calculations simple).

Table – I (calculations for ANOVA):

Technician K=4	**Errors (n = 5days)**					**Row total** T_i	$\frac{T_i^2}{n_i}$	$\sum X_{ij}^2$
	1	**2**	**3**	**4**	**5**			
I	-4	4	0	-2	1	-1	$\frac{(-1)^2}{5} = 0.2$	37
II	4	-1	2	0	4	9	$\frac{(9)^2}{5} = 16.2$	37
III	0	2	-3	5	1	5	$\frac{(5)^2}{5} = 5$	39
IV	-1	2	-2	0	1	0	$\frac{(0)^2}{5} = 0$	10
	C.F = $\frac{T^2}{N} = \frac{(13)^2}{20} = 8.45$					**Grand total** $T = \sum T_i = 13$	$\sum \frac{T_i^2}{n_i} = 21.4$	$\sum\sum X_{ij}^2 =$ 123

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 123 - 8.45 = 114.55$

Sum of squares between samples (SSB) = $\sum \frac{T_i^2}{n_i}$ - Correction Factor(C.F) = $21.4 - 8.45$

= 12.95

Mean square between samples (MSB) = $\frac{\text{Sum of squares between samples (SSB)}}{D.f \to \text{K}-1} = \frac{12.95}{4-1} = \frac{12.95}{3}$

$= 4.32$

Sum of squares within samples (SSW) = Total sum of squares (T.S.S) - Sum of squares between samples (SSB)

= 114.55 – 12.95 = 101.6

Mean square within samples (MSW) = $\frac{\text{Sum of squares within samples (SSW)}}{D.f \to \text{N}-\text{K}} = \frac{101.6}{20-4} = \frac{101.6}{16}$

$= 6.35$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Between samples	SSB = 12.95	$v_1 = 4 - 1 = 3$	$\mathrm{MSB} = \dfrac{\mathrm{SSB}}{\mathrm{K}-1} = 4.32$	$\mathrm{F_C} = \dfrac{\mathrm{MSW}}{\mathrm{MSB}} = 1.47$
Within samples	SSW = 101.6	$v_2 = 20 - 4 = 16$	$\mathrm{MSW} = \dfrac{\mathrm{SSW}}{\mathrm{N}-\mathrm{K}} = 6.35$	

∴ Calculated value of F is 1.47

Tabulated value of F for (16, 3) $(\mathrm{MSW} > MSB)$ at 1% l.o.s is 5.29.

∵ $\mathrm{F_C} < \mathrm{F_T}$ Null hypothesis H_0 is accepted.

HOME WORK:

1. **There are 3 typists working in an office. The times (in minutes) they spend for the tea break in addition to the allowed lunch break are observed and noted below:**

A:	**25**	**18**	**30**	**32**	**35**	**37**	**19**	**-**	**-**	**-**
B:	**24**	**22**	**26**	**28**	**30**	**32**	**28**	**26**	**-**	**-**
C:	**28**	**20**	**27**	**19**	**29**	**35**	**30**	**23**	**27**	**32**

Can the difference in average times that the 3 typists spend for the tea break be attributed to chance variation. ($\mathrm{F_T}$ at 5% l.o.s D.f (22, 2) is 19.45) **ANS:** $\mathbf{F_C = 11.62}$

2. **Three different machines are used for a production. On the basis of outputs, set up one-way ANOVA table and test whether the machines are equally effective.**

outputs		
Machine I	**Machine II**	**Machine III**
10	**9**	**20**
15	**7**	**16**
11	**5**	**10**
10	**6**	**14**

Given that the value of F at 5% l.o.s for (2, 9) d.f is 4.26. **ANS:** $\mathbf{F_C = 5.95}$

Two Way Classifications:

Randomised block design (R.B.D):

This design may be defined as one in which the experimental material is divided into blocks or groups in such a way that the units in each block are homogenous. The treatments are assigned at random to the units in each block. The layout of a R.B.D using three varieties of wheat in four blocks is given below.

Blocks			
I	II	III	IV
W_2	W_1	W_3	W_2
W_1	W_3	W_2	W_1
W_3	W_2	W_1	W_3

Here W's represent the random assignment of these varieties in each block. If the varieties represent a single criterian of classification that data can be recorded in a table shown below:

	Varieties		
Blocks	W_1	W_2	W_3
I	x_{11}	x_{12}	x_{13}
II	x_{21}	x_{22}	x_{23}
III	x_{31}	x_{32}	x_{33}
IV	x_{41}	x_{42}	x_{43}

Where x's denote observed data obtained from various blocks using different varieties of wheat. The ANOVA procedure for this design is the nearly same as discussed in one way classification.

ANOVA – Table:

Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F-Ratio**
Column treatments	SSC	$v_1 = c - 1$	$\mathrm{MSC} = \dfrac{\mathrm{SSC}}{\mathrm{c} - 1}$	$\mathrm{F_C} = \dfrac{\mathrm{MSC}}{\mathrm{MSE}}$
Row treatments	SSR	$v_2 = r - 1$	$\mathrm{MSR} = \dfrac{\mathrm{SSR}}{\mathrm{r} - 1}$	$\mathrm{F_R} = \dfrac{\mathrm{MSR}}{\mathrm{MSE}}$
Error (or) Residual	SSE	$(r-1)(c-1)$	$\mathrm{MSE} = \dfrac{\mathrm{SSE}}{(r-1)(c-1)}$	

Where,

SSC = **Sum of squares between blocks (columns)**

SSR = **Sum of squares between varieties (rows)**

MSC = **Mean square between blocks (columns)**

MSR = **Mean square between varieties (rows)**

SSE = **Residual Sum of squares = T.S.S – (S.S.C + S.S.R)**

MSE = **Mean square error** $\mathrm{MSE} = \frac{\mathrm{SSE}}{(r-1)(c-1)}$

Advantages of R.B.D:

1. This design is more efficient than C.R.D.
2. It is more accurate than C.R.D because it reduces experimental error.
3. It is more flexible. Any number of treatments and any number of replications may be used.
4. The analysis of design is simple, easily adaptable.

Disadvantages:

1. If the number of treatments is very large, then the size of the blocks will increase and this may introduce heterogeneity within blocks.
2. If the interactions are large, the experiment may yield misleading results.

Problems:

1. **Three varieties A, B, C of a crop are tested in a randomized block design with four replications. The plot yields in pounds are as follows.**

A(6)	**C(5)**	**A(8)**	**B(9)**
C(8)	**A(4)**	**B(6)**	**C(9)**
B(7)	**B(6)**	**C(10)**	**A(6)**

Solution:

Let us reframe the table by applying the principles of experimental design.

Varieties K=3	**Yield of the plots**				Total
	Number of Blocks				
	1	**2**	**3**	**4**	
A	6	8	4	6	24(R_1)
B	9	6	7	6	28(R_2)
C	5	8	9	10	32(R_3)
Total	21(C_1)	15(C_2)	24(C_3)	24(C_4)	84(T)

N -Total number of items in the given data = 12.

c - Number of columns = 4.

r - Number of rows = 3.

Null Hypothesis H_{01}: There is no significant difference between **Yields (columns) and**

H_{02}: There is no significant difference between **Varieties (rows)**

Table – I (calculations for ANOVA):

Treatments K=3	Yield of the plots / Number of Blocks 1	2	3	4	Row total R_j	$\frac{R_j^2}{n_j}$	$\sum_j X_{ij}^2$
A	6	8	4	6	24	$\frac{24^2}{4} = 144$	152
B	9	6	7	6	28	$\frac{28^2}{4} = 196$	202
C	5	8	9	10	32	$\frac{32^2}{4} = 256$	270
Column total C_i	21	15	24	24	Grand total G = 84	$\sum \frac{R_j^2}{n_j} = 596$	*
$\frac{C_i^2}{n_i}$	$\frac{21^2}{3} = 147$	$\frac{15^2}{3} = 75$	$\frac{24^2}{3} = 192$	$\frac{24^2}{3} = 192$	$\sum \frac{C_j^2}{n_i} = 606$	C.F = $\frac{G^2}{N} = \frac{84^2}{12} = 588$	*
$\sum_i X_{ij}^2$	142	164	146	172	*	*	$\sum\sum X_{ij}^2$ =624

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 624 - 588 = 36$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n_i}$ - Correction Factor (C.F) = $606 - 588 = 18$

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to c-1} = \frac{18}{4-1} = \frac{18}{3} = 6$

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n_j}$ - Correction Factor (C.F) = $596 - 588 = 8$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to r-1} = \frac{8}{3-1} = \frac{8}{2} = 4$

SSE = Residual Sum of squares = T.S.S – (S.S.C + S.S.R) = 36 – (18+8) = 36 – 26 = 10

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = \frac{10}{(3-1)(4-1)} = \frac{10}{(2)(3)} = \frac{10}{6} = 1.667$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Column treatments	SSC = 18	$v_1 = 4 - 1 = 3$	$\text{MSC} = \frac{\text{SSC}}{c-1} = 6$	$F_C = \frac{\text{MSC}}{\text{MSE}} = 3.6$
Row treatments	SSR = 8	$v_2 = 3 - 1 = 2$	$\text{MSR} = \frac{\text{SSR}}{r-1} = 4$	$F_R = \frac{\text{MSR}}{\text{MSE}} = 2.4$
Error (or) Residual	SSE = 10	$(r-1)(c-1) = 6$	$\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = 1.667$	*

Conclusion:

(i) For Yields:

Tabulated value of F for (3, 6) at 5% l.o.s is 4.76.

Calculated value of F is $F_C = 3.6$

$\because F_C < F_T$ Null hypothesis H_{01} is accepted.

(ii) For Varieties:

Tabulated value of F for (2, 6) at 5% l.o.s is 5.14.

Calculated value of F is $F_R = 2.4$

$\because F_R < F_T$ Null hypothesis H_{02} is accepted.

2. **The yield of four strains of a particular variety of wheat was planted in five randomized blocks in Kgs per plot is given below.**

Strains	Blocks				
	1	2	3	4	5
A	32	34	34	35	36
B	33	33	36	37	34
C	30	35	35	32	35
D	29	22	30	28	28

Test for difference between blocks and difference between strains.

Solution:

N -Total number of items in the given data = 20.

c - Number of columns = 5.

r - Number of rows = 4.

Null Hypothesis H_{01}: There is no significant difference between **Blocks (columns) and**

H_{02}: There is no significant difference between **Strains (rows)**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 30 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Strains K=4	Blocks					Row total R_j	$\frac{R_j^2}{n_j}$	$\sum_j X_{ij}^2$
	1	2	3	4	5			
A	2	4	4	5	6	21	$\frac{21^2}{5} = 88.2$	97
B	3	3	6	7	4	23	$\frac{23^2}{5} = 105.8$	119
C	0	5	5	2	5	17	$\frac{17^2}{5} = 57.8$	79
D	-1	-8	0	-2	-2	-13	$\frac{(-13)^2}{5} = 33.8$	73
Column total C_i	4	4	15	12	13	Grand total G = 48	$\sum \frac{R_j^2}{n_j} = 285.6$	*
$\frac{C_i^2}{n_i}$	$\frac{4^2}{4} = 4$	$\frac{4^2}{4} = 4$	$\frac{15^2}{4} = 56.25$	$\frac{12^2}{4} = 36$	$\frac{13^2}{4} = 42.25$	$\sum \frac{C_j^2}{n_i} = 142.5$	C.F = $\frac{G^2}{N} = \frac{48^2}{20} = 115.2$	*
$\sum_i X_{ij}^2$	14	114	77	82	81	*	*	$\sum\sum X_{ij}^2$ =368

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 368 - 115.2 = 252.8$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n_i}$ - Correction Factor (C.F) = $142.5 - 115.2$ = 27.3

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to \text{c}-1} = \frac{27.3}{5-1} = \frac{27.3}{4}$ =6 .825

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n_j}$ - Correction Factor (C.F) = $285.6 - 115.2 = 170.4$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to \text{r}-1} = \frac{170.4}{4-1} = \frac{170.4}{3} = 56.8$

SSE = Residual Sum of squares = TSS – (S.S.C + S.S.R) = 252.8 – (27.3+170.4) = 368 – 197.7 = 55.1

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = \frac{55.1}{(4-1)(5-1)} = \frac{55.1}{(3)(4)} = \frac{55.1}{12} = 4.59$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Column treatments	SSC = 27.3	$\upsilon_1 = 5 - 1 = 4$	$\mathrm{MSC} = \dfrac{\mathrm{SSC}}{\mathrm{c} - 1} = 6.825$	$\mathrm{F_C} = \dfrac{\mathrm{MSC}}{\mathrm{MSE}} = 1.486$
Row treatments	SSR = 170.4	$\upsilon_2 = 4 - 1 = 3$	$\mathrm{MSR} = \dfrac{\mathrm{SSR}}{\mathrm{r} - 1} = 56.8$	$\mathrm{F_R} = \dfrac{\mathrm{MSR}}{\mathrm{MSE}} = 12.37$
Error (or) Residual	SSE = 55.1	$(r-1)(c-1) = 12$	$\mathrm{MSE} = \dfrac{\mathrm{SSE}}{(r-1)(c-1)} = 4.59$	*

Conclusion:

(i) For Blocks:

Tabulated value of F for (4, 12) at 5% l.o.s is 5.91.

Calculated value of F is $\mathrm{F_C} = 1.486$

$\because \mathrm{F_C} < \mathrm{F_T}$ Null hypothesis H_{01} is accepted.

(ii) For Strains:

Tabulated value of F for (3, 12) at 5% l.o.s is 8.74.

Calculated value of F is $\mathrm{F_R} = 12.37$

$\because \mathrm{F_R} > \mathrm{F_T}$ Null hypothesis H_{02} is rejected.

3. The following data represent the number of units of production per day turned out by four randomly chosen operators using three milling machines.

Operators	Machines		
	M I	**M II**	**M III**
1	**150**	**151**	**156**
2	**147**	**159**	**155**
3	**141**	**146**	**153**
4	**154**	**152**	**159**

Perform analysis of variance and test hypothesis

(a) That the machines are not significantly different.

(b) That the operators are not significantly different at 5% level.

Solution:

N -Total number of items in the given data = 12.

c - Number of columns = 3.

r - Number of rows = 4.

Null Hypothesis H_{01}: There is no significant difference between **Machines (columns) and**

H_{02}: There is no significant difference between **Operators (rows)**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 150 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Operators K=4	Blocks			Row total R_j	$\frac{R_j^2}{n_j}$	$\sum_j X_{ij}^2$
	M I	M II	M III			
1	0	1	6	7	$\frac{7^2}{3} = 16.33$	37
2	-3	9	5	11	$\frac{11^2}{3} = 40.33$	115
3	-9	-4	3	-10	$\frac{(-10)^2}{3} = 33.33$	106
4	4	2	9	15	$\frac{15^2}{3} = 75$	101
Column total C_i	-8	8	23	**Grand total** G = 23	$\sum \frac{R_j^2}{n_j} = 164.99$	*
$\frac{C_i^2}{n_i}$	$\frac{(-8)^2}{4} = 16$	$\frac{8^2}{4} = 16$	$\frac{23^2}{4} = 132.25$	$\sum \frac{C_j^2}{n_i} = 164.25$	C.F = $\frac{G^2}{N} = \frac{23^2}{12} = 44.08$	*
$\sum_i X_{ij}^2$	106	102	151	*	*	$\sum\sum X_{ij}^2$ =359

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}{}^2)$ - Correction Factor(C.F)

$= 359 - 44.08 = 314.92$

Sum of squares between columns (SSC) = $\sum \frac{C_i{}^2}{n_i}$ - Correction Factor (C.F) = $164.25 - 44.08$

=120.17

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to \text{c}-1} = \frac{120.17}{3-1} = \frac{120.17}{2}$

= 60 .085

Sum of squares between rows (SSR) = $\sum \frac{R_j{}^2}{n_j}$ - Correction Factor (C.F) $= 164.99 - 44.08$

$= 120.91$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to \text{r}-1} = \frac{120.91}{4-1} = \frac{120.91}{3} = 40.30$

SSE = Residual Sum of squares = TSS – (SSC + SSR) = 314.92 – (120.17+120.91)

= 314.92 – 361.25 = 73.84

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = \frac{73.84}{(4-1)(3-1)} = \frac{73.84}{(3)(2)} = \frac{73.84}{6} = 12.31$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Column treatments	SSC = 120.17	$v_1 = 4 - 1 = 3$	$\text{MSC} = \frac{\text{SSC}}{\text{c} - 1} = 60 .085$	$\text{F}_\text{C} = \frac{\text{MSC}}{\text{MSE}} = 4.88$
Row treatments	SSR = 120.91	$v_2 = 3 - 1 = 2$	$\text{MSR} = \frac{\text{SSR}}{\text{r} - 1} = 40.30$	$\text{F}_\text{R} = \frac{\text{MSR}}{\text{MSE}} = 3.27$
Error (or) Residual	SSE = 73.84	$(r-1)(c-1) = 6$	$\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = 12.31$	*

Conclusion:

(i) For Machines:

Tabulated value of F for (3, 6) at 5% l.o.s is 4.76.

Calculated value of F is $\text{F}_\text{C} = 4.88$

$\because F_C > F_T$ Null hypothesis H_{01} is rejected.

(ii) For Operators:

Tabulated value of F for (2, 6) at 5% l.o.s is 5.14.

Calculated value of F is $F_R = 3.27$

$\because F_R < F_T$ Null hypothesis H_{02} is accepted.

4. Five doctors each test treatments for a certain disease and observe the number of days each take to recover. The results are as follows: (recovery time in days)

Doctor	Treatments				
	1	2	3	4	5
A	10	14	23	19	20
B	11	15	24	17	21
C	9	12	20	16	19
D	8	13	17	17	20
E	12	15	19	15	22

Discuss the difference between (a) doctors and (b) treatments

Solution:

N -Total number of items in the given data = 25.

c - Number of columns = 5.

r - Number of rows = 5.

Null Hypothesis H_{01}: There is no significant difference between **Treatments (columns) and**

H_{02}: There is no significant difference between **Doctors (rows)**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 16 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Doctors K=5	Treatments					Row total R_j	$\frac{R_j^2}{n_j}$	$\sum_j X_{ij}^2$
	1	2	3	4	5			
A	-6	-2	7	3	4	6	$\frac{6^2}{5} = 7.2$	114
B	-5	-1	8	1	5	8	$\frac{8^2}{5} = 12.8$	116
C	-7	-4	4	0	3	-4	$\frac{-4^2}{5} = 3.2$	90
D	-8	-3	1	1	4	-5	$\frac{(-5)^2}{5} = 5$	91
E	-4	-1	3	-1	6	3	$\frac{3^2}{5} = 1.8$	63
Column total C_i	-30	-11	23	4	22	Grand total G = 8	$\sum \frac{R_j^2}{n_j} = 30$	*
$\frac{C_i^2}{n_i}$	$\frac{(-30)^2}{5} = 180$	$\frac{(-11)^2}{5} = 24.2$	$\frac{23^2}{5} = 105.8$	$\frac{4^2}{5} = 3.2$	$\frac{22^2}{5} = 96.8$	$\sum \frac{C_j^2}{n_i} = 410$	C.F $= \frac{G^2}{N} = \frac{8^2}{25} = 2.56$	*
$\sum_i X_{ij}^2$	190	31	139	12	102	*	*	$\sum\sum X_{ij}^2$ =474

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$= 474 - 2.56 = 471.44$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n_i}$ - Correction Factor (C.F) = $410 - 2.56$ = 407.44

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to \text{c}-1} = \frac{407.44}{5-1} = \frac{407.44}{4}$

=101.86

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n_j}$ - Correction Factor (C.F) = $30 - 2.56 = 27.44$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to \text{r}-1} = \frac{27.44}{5-1} = \frac{170.4}{4} = 6.86$

SSE = Residual Sum of squares = TSS – (SSC + SSR) = 471.44 – (407.44+27.44)

= 471.44 – 197.7 = 36.56

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = \frac{36.56}{(5-1)(5-1)} = \frac{36.56}{(4)(4)} = \frac{36.56}{16} = 2.285$

Table-II (ANOVA table):

Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F-Ratio**
Column treatments	SSC = 407.44	$v_1 = 5 - 1 = 4$	$\text{MSC} = \frac{\text{SSC}}{\text{c}-1} = 101.86$	$\text{F}_\text{C} = \frac{\text{MSC}}{\text{MSE}} = 44.58$
Row treatments	SSR = 27.44	$v_2 = 5 - 1 = 4$	$\text{MSR} = \frac{\text{SSR}}{\text{r}-1} = 6.86$	$\text{F}_\text{R} = \frac{\text{MSR}}{\text{MSE}} = 3.00$
Error (or) Residual	SSE = 36.56	$(r-1)(c-1) = 16$	$\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = 2.285$	*

Conclusion:

(i) For Treatments:

Tabulated value of F for (4, 16) at 5% l.o.s is 3.01.

Calculated value of F is $\text{F}_\text{C} = 44.58$

$\because \text{F}_\text{C} > \text{F}_\text{T}$ Null hypothesis H_{01} is rejected.

(ii) For Doctors:

Tabulated value of F for (4, 16) at 5% l.o.s is 3.01.

Calculated value of F is $\text{F}_\text{R} = 3.00$

$\because \text{F}_\text{R} < \text{F}_\text{T}$ Null hypothesis H_{02} is accepted.

5. **A company appoints 4 salesman A, B, C, D & observes their sales in 3 seasons: summer, winter, monsoon. The figures (in lakhs) are given in the following table.**

Season	**Salesmen**			
	A	**B**	**C**	**D**
Summer	**45**	**40**	**38**	**37**
Winter	**43**	**41**	**45**	**38**
Monsoon	**39**	**39**	**41**	**41**

Carry out an analysis of variance.

Solution:

N -Total number of items in the given data = 12.

c - Number of columns = 4.

r - Number of rows = 3.

Null Hypothesis H_{01}: There is no significant difference between **Salesmen (columns) and**

H_{02}: There is no significant difference between **Seasons (rows)**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 40 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Season K=3	Salesmen				Row total R_j	$\frac{R_j^2}{n_j}$	$\sum_j X_{ij}^2$
	A	B	C	D			
Summer	5	0	-2	-3	0	$\frac{0^2}{4} = 0$	38
Winter	3	1	5	-2	7	$\frac{7^2}{4} = 12.25$	39
Monsoon	-1	-1	1	1	0	$\frac{0^2}{4} = 0$	4
Column total C_i	7	0	4	-4	Grand total G = 7	$\sum \frac{R_j^2}{n_j} = 12.25$	*
$\frac{C_i^2}{n_i}$	$\frac{7^2}{3} = 16.33$	$\frac{0^2}{3} = 0$	$\frac{4^2}{3} = 5.33$	$\frac{(-4)^2}{3} = 5.33$	$\sum \frac{C_j^2}{n_i} = 26.99$	C.F = $\frac{G^2}{N} = \frac{7^2}{12} = 4.08$	*
$\sum_i X_{ij}^2$	35	2	30	14	*	*	$\sum\sum X_{ij}^2$ =81

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

=$81 - 4.08 = 76.92$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n_i}$ - Correction Factor (C.F) = $26.99 - 4.08 = 22.91$

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to c-1} = \frac{22.91}{4-1} = \frac{22.91}{3}$ =7.64

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n_j}$ - Correction Factor (C.F) = $12.25 - 4.08 = 8.17$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to r-1} = \frac{8.17}{3-1} = \frac{8.17}{2} = 4.085$

SSE = Residual Sum of squares = T.S.S – (S.S.C + S.S.R) = 76.92 – (22.91+8.17) = 76.92 – 31.08

= 45.84

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = \frac{45.84}{(3-1)(4-1)} = \frac{45.84}{(2)(3)} = \frac{45.84}{6} = 7.64$

Table-II (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Column treatments	SSC = 22.91	$v_1 = 4 - 1 = 3$	$\text{MSC} = \frac{\text{SSC}}{c-1} = 7.64$	$F_C = \frac{\text{MSC}}{\text{MSE}} = 1$
Row treatments	SSR =8.17	$v_2 = 3 - 1 = 2$	$\text{MSR} = \frac{\text{SSR}}{r-1} = 4.085$	$F_R = \frac{\text{MSE}}{\text{MSR}} = 1.87$
Error (or) Residual	SSE = 10	$(r-1)(c-1) = 6$	$\text{MSE} = \frac{\text{SSE}}{(r-1)(c-1)} = 7.64$	*

Conclusion:

(i) For Salesmen:

Tabulated value of F for (3, 6) at 5% l.o.s is 4.76.

Calculated value of F is $F_C = 1$

$\because F_C < F_T$ Null hypothesis H_{01} is accepted.

(ii) For Seasons:

Tabulated value of F for (2, 6) at 5% l.o.s is 5.14.

Calculated value of F is $F_R = 1.87$

$\because F_R < F_T$ Null hypothesis H_{02} is accepted.

HOME WORK:

1. **Four varieties A, B, C, D of a fertilizer are tested in a RBD with four replications. The plot yields in pounds are as follows.**

A(12)	D(20)	C(16)	B(10)
D(18)	A(14)	B(11)	C(14)
B(12)	C(15)	D(19)	A(13)
C(16)	B(11)	A(15)	D(20)

Analyze the experimental yield. (A.U APR'06) (ANS: $F_C = 1.74, F_{T0.05,(9,3)} = 8.82$

$F_R = 41.51, F_{T0.05,(3,9)} = 3.86$**)**

2. **Four varieties A, B, C, D of a fertilizer are tested in a RBD with four replications. The plot yields in pounds are as follows.**

B(90)	E(80)	C(134)	A(112)	D(92)
E(85)	D(84)	B(70)	C(141)	A(82)
C(110)	A(90)	D(87)	B(84)	E(69)
A(81)	C(125)	E(85)	D(76)	B(72)
D(82)	B(60)	A(94)	E(85)	C(88)

Carry out an analysis of variance, what inference can you draw from the data given.

(ANS: $F_C = 1.742, F_{T0.05,(4,16)} = 3.01$ **(A.U JUN'06)**

$F_R = 10.542, F_{T0.05,(4,16)} = 3.01$**)**

3. **The following data represent the number of units of production per day turned out by 5 different workers using 4 different types of machines.**

Workers	Machine type			
	A	B	C	D
1	44	38	47	36
2	46	40	52	43
3	34	36	44	32
4	43	38	46	33
5	38	42	49	39

(a) Test whether the mean production is the same for the different machine types.

(b) Test whether the 5 men differ with mean productivity.

(ANS: $F_C = 18.38, F_{T0.05,(3,12)} = 3.49$

$F_R = 6.574, F_{T0.05,(2,12)} = 3.26$**)**

4. **A tea company appoints 4 salesman A, B, C, D & observes their sales in 3 seasons: summer, winter, monsoon. The figures (in lakhs) are given in the following table.**

Season	Salesmen				Season's Total
	A	B	C	D	
Summer	36	36	21	35	128
Winter	28	29	31	32	120
Monsoon	26	28	29	29	112
Salesmen's Total	90	93	81	96	360

(i) Do the salesmen significantly differ in performance?

(ii) Is there significant difference between seasons?

(ANS: $F_C = 1.619, F_{T0.05,(6,3)} = 8.94$

$F_R = 1.417, F_{T0.05,(6,2)} = 19.35$**)**

5. **To study the performance of three detergents and three different water temperature, the following "whiteness" readings were obtained:**

Water temperature	Detergent A	Detergent B	Detergent C
Cold water	57	55	67
Warm water	49	52	68
Hot water	54	46	58

Perform a two way analysis of variance, using 5% l.o.s.

(ANS: $F_C = 9.86, F_{T0.05,(2,4)} = 6.94$

$F_R = 2.38, F_{T0.05,(2,4)} = 6.94$**)**

Latin square design:

A Latin square is a square arrangement of m-rows and m-columns such that each symbol appears once in each row and column. For example,

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

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

Suppose there are three factors i.e., two types of feed each of five different strengths to be tried on 25 animals of five different litters (young's of some animals) five for each litter. We name the litters as A, B, C, D, E and set the animals in such a manner as to form a Latin square containing five rows and columns so that animal of each litter appears once in each row and each column. Each of the five rows is related with one of the strengths of the first feed say F_1 and each column with one of the five strengths of second feed say F_2. Thus a possible arrangement can be

	Second feed F_2					
	*	1	2	3	4	5
First feed F_1	1	A	B	C	D	E
	2	B	E	A	C	D
	3	C	D	B	E	A
	4	D	C	E	A	B
	5	E	A	D	B	C

The observation x_{ij} in the i^{th} row and j^{th} column may be the gain in weight of the animal in the position. The Latin square arrangement design would eliminate the effect of the litters and one feed to observe the effect of other feed.

The above arrangement is called Latin square arrangement.

NOTE:

1. If there are n rows and n columns then the order of the Latin square is n.
2. If the first row and first column has alphabets in alphabetical order, then it is called standard square.
3. Latin square design is very useful in controlling two sources of variation and at the same time reduces the number of treatment combinations.
4. The assumption made in Latin square design is that interactions between treatments and row and column groupings are non-existent.

Advantages:

1. Latin square arrangement design controls more of the variation than the CRBD with a two way stratification.
2. The analysis is simple.
3. Even with missing data the analysis remains relatively simple.

Randomized block design (RBD) & Latin square design (LSD) comparison:

RBD	**LSD**
1. The RBD is superior to LSD in many ways.	**1.** The LSD can be applied only in special occasions.
2. RBD can be applied for square or rectangular field.	**2.** LSD can be applied only for square field.

3. RBD is available for wide range of treatments (2 to 24)	**3.** LSD is suitable when the number of treatments is 5 to 10.
4. RBD controls the effect of only one extraneous variable.	**4.** LSD controls the effect of two extraneous variables.
5. In RBD, there is no restrictions on the number of replications and number of treatments.	**5.** In LSD, The number of replications of each treatment is equal to the number of treatments.
6. The ANOVA is more flexible in RBD.	**6.** The ANOVA is not more flexible in LSD.
7. In RBD there are no restrictions for comparison of replications & treatments.	**7.** In LSD, for a comparison of a smaller number of treatments, the number of replications is inadequate while for a large number of treatments, the number has to be unduly increased.

The ANOVA procedure for this design is the nearly same as discussed in Two way classification.

ANOVA – Table:

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Columns	SSC	$n-1$	$\mathrm{MSC} = \dfrac{\mathrm{SSC}}{n-1}$	$\mathrm{F_C} = \dfrac{\mathrm{MSC}}{\mathrm{MSE}}$
Rows	SSR	$n-1$	$\mathrm{MSR} = \dfrac{\mathrm{SSR}}{n-1}$	$\mathrm{F_R} = \dfrac{\mathrm{MSR}}{\mathrm{MSE}}$
Treatments	SST	$n-1$	$\mathrm{MST} = \dfrac{\mathrm{SST}}{n-1}$	$\mathrm{F_T} = \dfrac{\mathrm{MST}}{\mathrm{MSE}}$
Error (or) Residual	SSE	$(n-1)(n-2)$	$\mathrm{MSE} = \dfrac{\mathrm{SSE}}{(n-1)(n-2)}$	*

Where,

SSC = **Sum of squares between columns**

SSR = **Sum of squares between rows**

SST = **Sum of squares between rows**

MSC = **Mean square between treatments**

MSR = **Mean square between rows**

MST = **Mean square between treatments**

SSE = **Residual Sum of squares = T.S.S – (S.S.C + S.S.R+S.S.T)**

MSE = **Mean square error** $\mathrm{MSE} = \frac{\mathrm{SSE}}{(n-1)(n-2)}$

Problems:

1. Analyze the following results of a Latin square experiments:

	1	2	3	4
1	**A(12)**	**D(20)**	**C(16)**	**B(10)**
2	**D(18)**	**A(14)**	**B(11)**	**C(14)**
3	**B(12)**	**C(15)**	**D(19)**	**A(13)**
4	**C(16)**	**B(11)**	**A(15)**	**D(20)**

The letter A, B, C, D denotes the treatments and the figures in brackets denote the observations

Solution:

N -Total number of items in the given data = 16.

c - Number of columns = 4.

r - Number of rows = 4.

Null Hypothesis H_{01}: There is no significant difference between **columns.**

H_{02}: There is no significant difference between **rows** and

H_{03}: There is no significant difference between **treatments.**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 15 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Rows K=4	Columns				Row total R_j	$\frac{R_j^2}{n}$	$\sum_j X_{ij}^2$
	C_1	C_2	C_3	C_4			
R_1	A(-3)	D(5)	C(1)	B(-5)	-2	$\frac{(-2)^2}{4} = 1$	60
R_2	D(3)	A(-1)	B(-4)	C(-1)	-3	$\frac{(-3)^2}{4} = 2.25$	27
R_3	B(-3)	C(0)	D(4)	A(-2)	-1	$\frac{(-1)^2}{4} = 0.25$	29
R_4	C(1)	B(-4)	A(0)	D(5)	2	$\frac{2^2}{4} = 1$	42
Column total C_i	-2	0	1	-3	Grand total G = -4	$\sum \frac{R_j^2}{n} = 4.50$	*
$\frac{C_i^2}{n}$	$\frac{(-2)^2}{4} = 1$	$\frac{0^2}{4} = 0$	$\frac{1^2}{4} = 0.25$	$\frac{(-3)^2}{4} = 2.25$	$\sum \frac{C_j^2}{n} = 3.5$	C.F = $\frac{G^2}{N} = \frac{(-4)^2}{16} = 1$	*
$\sum_i X_{ij}^2$	28	42	33	55	*	*	$\sum\sum X_{ij}^2$ =158

Table – II (To find SST):

Alphabets (in order)	x_{ij} (Main entries of table – I in order)					
	1	2	3	4	Row total T_i	$\frac{T_i^2}{n=4}$
A	-3	-1	0	-2	-6	**9**
B	-3	-4	-4	-5	16	**64**
C	1	0	1	-1	1	**0.25**
D	3	5	4	5	17	**72.25**
						$\sum \frac{T_i^2}{4}$ **=145.5**

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$=158 - 1 = 157$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n}$ - Correction Factor (C.F) = $3.5 - 1$ = 2.5

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \to n-1} = \frac{2.5}{4-1} = \frac{2.5}{3}$ =0.833

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n}$ - Correction Factor (C.F) = $4.5 - 1 = 3.5$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \to n-1} = \frac{3.5}{4-1} = \frac{3.5}{3} = 1.167$

Sum of squares between columns (SST) = $\sum \frac{T_i^2}{n}$ - Correction Factor (C.F) = $145.5 - 1$ = 144.5

Mean square between columns (MST) = $\frac{\text{Sum of squares between samples (SST)}}{D.f \to n-1} = \frac{144.5}{4-1} = \frac{144.5}{3}$

$= 48.167$

SSE = Residual Sum of squares = T.S.S – (SSC + SSR+SST) = 157 – (2.5+3.5+ 144.5)

= 157 – 150.5 = 6.5.

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(n-1)(n-2)} = \frac{6.5}{(4-1)(4-2)} = \frac{6.5}{(3)(2)} = \frac{6.5}{6} = 1.08$

Table-III (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Columns	SSC = 2.5	$n-1$ = 3	$\text{MSC} = \frac{\text{SSC}}{n-1}$ = 0.87	$\text{F}_\text{C} = \frac{\text{MSE}}{\text{MSC}}$ = 1.24
Rows	SSR = 3.5	$n-1$ = 3	$\text{MSR} = \frac{\text{SSR}}{n-1}$ = 1.167	$\text{F}_\text{R} = \frac{\text{MSR}}{\text{MSE}}$ = 1.081
Treatments	SST = 144.5	$n-1$ = 3	$\text{MST} = \frac{\text{SST}}{n-1}$ = 48.17	$\text{F}_\text{T} = \frac{\text{MST}}{\text{MSE}}$ = 44.60
Error (or) Residual	SSE = 6.5	$(n-1)(n-2)$ = 6	$\text{MSE} = \frac{\text{SSE}}{(n-1)(n-2)}$ = 1.08	*

Conclusion:

(i) For Columns:

Tabulated value of F for (6, 3) at 5% l.o.s is 8.94.

Calculated value of F is $\text{F}_\text{C} = 1.24$

$\because \text{F}_\text{C} < \text{F}_\text{Tab}$ Null hypothesis H_{01} is accepted.

(ii) For Rows:

Tabulated value of F for (3, 6) at 5% l.o.s is 4.76.

Calculated value of F is $\text{F}_\text{R} = 1.081$

$\because \text{F}_\text{R} < \text{F}_\text{Tab}$ Null hypothesis H_{02} is accepted.

(iii) For Treatments:

Tabulated value of F for (3, 6) at 5% l.o.s is 4.76.

Calculated value of F is $\text{F}_\text{T} = 44.60$

$\because \text{F}_\text{T} > \text{F}_\text{Tab}$ Null hypothesis H_{03} is rejected.

2. **An agricultural experiment was conducted on Latin square to investigate the yield per acre of five different varieties of wheat when subjected to treatment of two types of fertilizers A and B each of five different strengths. The results are set out in a Latin square below in which letters corresponds to varieties, columns are treated with different strengths of fertilizers B.**

A16	**B10**	**C11**	**D9**	**E9**
E10	**C9**	**A14**	**B12**	**D11**
B15	**D8**	**E8**	**C10**	**A18**
D12	**E6**	**B13**	**A13**	**C12**
C13	**A11**	**D10**	**E7**	**B14**

Discuss the yield variations w. r. t each of the factors.

Solution:

N -Total number of items in the given data = 20.

c - Number of columns = 5.

r - Number of rows = 5.

Null Hypothesis H_{01}: There is no significant difference between **columns.**

H_{02}: There is no significant difference between **rows** and

H_{03}: There is no significant difference between **treatments.**

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 11 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Rows K=4	Columns					Row total R_j	$\frac{R_j^2}{n}$	$\sum_j X_{ij}^2$
	C_1	C_2	C_3	C_4	C_5			
R_1	A(5)	B(-1)	C(0)	D(-2)	E(-2)	0	$\frac{(0)^2}{5} = 0$	34
R_2	E(-1)	C(-2)	A(3)	B(1)	D(0)	1	$\frac{(1)^2}{5} = 0.20$	15
R_3	B(4)	D(-3)	E(-3)	C(-1)	A(7)	4	$\frac{(4)^2}{5} = 3.2$	84
R_4	D(1)	E(-5)	B(2)	A(2)	C(1)	1	$\frac{1^2}{5} = 0.20$	35
R_5	C(2)	A(0)	D(-1)	E(-4)	B(3)	0	$\frac{0^2}{5} = 0$	30
Column total C_i	11	-11	1	-4	9	Grand total G = 6	$\sum \frac{R_j^2}{n} = 3.6$	*
$\frac{C_i^2}{n}$	$\frac{(11)^2}{5} = 24.2$	$\frac{(-11)^2}{5} = 24.2$	$\frac{1^2}{5} = 0.20$	$\frac{(-4)^2}{5} = 3.2$	$\frac{(9)^2}{5} = 16.2$	$\sum \frac{C_j^2}{n} = 68$	C.F = $\frac{G^2}{N} = \frac{(6)^2}{25} = 1.44$	*
$\sum_i X_{ij}^2$	47	39	23	26	63	*	*	$\sum\sum X_{ij}^2$ =198

Table – II (To find SST):

Alphabets (in order)	x_{ij} (Main entries of table – I in order)						
	1	2	3	4	5	Row total T_i	$\frac{T_i^2}{n=5}$
A	5	0	3	2	7	17	57.8
B	4	-1	2	1	3	9	16.2
C	2	-2	0	-1	1	0	0
D	1	-3	-1	-2	0	-5	5
E	-1	-5	-3	-4	-2	-15	45
							$\sum \frac{T_i^2}{5}$ =124

Total sum of squares (T.S.S) = Sum of squares of all the items $(\sum\sum X_{ij}^2)$ - Correction Factor(C.F)

$=198 - 1.44 = 196.56$

Sum of squares between columns (SSC) = $\sum \frac{C_i^2}{n}$ - Correction Factor (C.F) = 68 − 1.44 = 66.56

Mean square between columns (MSC) = $\frac{\text{Sum of squares between samples (SSC)}}{D.f \rightarrow n-1} = \frac{66.56}{5-1} = \frac{66.56}{4}$

=16.64

Sum of squares between rows (SSR) = $\sum \frac{R_j^2}{n}$ - Correction Factor (C.F) = $3.6 - 1.44 = 2.16$

Mean square between rows (MSR) = $\frac{\text{Sum of squares within samples (SSR)}}{D.f \rightarrow n-1} = \frac{2.16}{5-1} = \frac{2.16}{4} = 0.54$

Sum of squares between columns (SST) = $\sum \frac{T_i^2}{n}$ - Correction Factor (C.F) = 124 − 1.44 = 122.56

Mean square between columns (MST) = $\frac{\text{Sum of squares between samples (SST)}}{D.f \to n-1} = \frac{122.56}{5-1} = \frac{122.56}{4}$

$= 30.64$

SSE = Residual Sum of squares = T.S.S – (SSC + SSR+SST) = 196.56 – (2.16+66.56+ 122.56)

= 196.56 – 191.28 = 5.28.

MSE = Mean square error $\text{MSE} = \frac{\text{SSE}}{(n-1)(n-2)} = \frac{5.28}{(5-1)(5-2)} = \frac{5.28}{(4)(3)} = \frac{5.28}{12} = 0.44$

Table-III (ANOVA table):

Source of variation	Sum of squares	Degree of freedom	Mean squares	F-Ratio
Columns	SSC = 66.56	$n-1$ = 4	$\text{MSC} = \frac{\text{SSC}}{n-1}$ = 16.64	$\text{F}_\text{C} = \frac{\text{MSC}}{\text{MSE}}$ = 37.81
Rows	SSR = 2.16	$n-1$ = 4	$\text{MSR} = \frac{\text{SSR}}{n-1}$ = 0.54	$\text{F}_\text{R} = \frac{\text{MSR}}{\text{MSE}}$ = 1.227
Treatments	SST = 122.56	$n-1$ = 4	$\text{MST} = \frac{\text{SST}}{n-1}$ = 30.64	$\text{F}_\text{T} = \frac{\text{MST}}{\text{MSE}}$ = 69.64
Error (or) Residual	SSE = 5.28	$(n-1)(n-2) = 12$	$\text{MSE} = \frac{\text{SSE}}{(n-1)(n-2)}$ = 0.44	*

Conclusion:

(i) For Columns:

Tabulated value of F for (4, 12) at 5% l.o.s is 3.26.

Calculated value of F is $\text{F}_\text{C} = 37.81$

$\because \text{F}_\text{C} > \text{F}_{\text{Tab}}$ Null hypothesis H_{01} is rejected.

(ii) For Rows:

Tabulated value of F for (4, 12) at 5% l.o.s is 3.26.

Calculated value of F is $\text{F}_\text{R} = 1.227$

$\because \text{F}_\text{R} < \text{F}_{\text{Tab}}$ Null hypothesis H_{02} is accepted.

(iii) For Treatments:

Tabulated value of F for (4, 12) at 5% l.o.s is 3.26.

Calculated value of F is $\text{F}_\text{T} = 69.64$

$\because \text{F}_\text{T} > \text{F}_{\text{Tab}}$ Null hypothesis H_{03} is rejected.

HOME WORK:

1. **Five varieties of paddy A, B, C, D, E are tried. The plan, the varieties shown in each plot and yields obtained in kg are given in the following table.**

B	**E**	**C**	**A**	**D**
95	**85**	**139**	**117**	**97**
E	**D**	**B**	**C**	**A**
90	**89**	**75**	**146**	**87**
C	**A**	**D**	**B**	**E**
116	**95**	**92**	**89**	**74**
A	**C**	**E**	**D**	**B**
85	**130**	**90**	**81**	**77**
D	**B**	**A**	**E**	**C**
87	**65**	**99**	**89**	**93**

Test whether there is a significant difference between rows and columns at 5% l.o.s. (ANS: F_{Tab} for (4, 12) at 5% l.o.s is 3.26 ($\text{F}_\text{C} = 1.374, \text{F}_\text{R} = 1.528, \text{F}_\text{T} = 7.276$**)**

2. **Five varieties of wheat A, B, C, D and E were studied. The plan, the varieties shown in each plot and yields obtained in kg are given in the following table.**

Carry out an ANOVA. What inference can you draw from the given data.

A90	**E80**	**C134**	**A112**	**D92**
E85	**D84**	**E70**	**C141**	**A82**
C110	**A90**	**D87**	**B84**	**E69**
A81	**C125**	**E85**	**D76**	**B72**
D82	**B60**	**A94**	**E85**	**C88**

(ANS: F_{Tab} for (4, 12) at 5% l.o.s is 3.26 ($F_C = 2.46, F_R = 2.64, F_T = 14.90$**)**

2^2-Factorial Design:

In the 2^2-Factorial Design we will have 2 factors each at two levels (0, 1) (say). Then there are 2×2 i.e., 4 treatment combinations in all. *This 2^2-Factorial Design can be performed in the form of* **CRD, RBD, LSD**. By Yate's notation "**K**" and "**P**" denote the 2 factors under study and 'k' and 'p' denote one of the 2 levels of each corresponding factors and this is called second level. The first level of "**K**" and "**P**" is generally expressed by the absence of the corresponding letter in the treatment combinations.

The four treatment combinations are,

$\boldsymbol{k_0 p_0 (or) 1}$ **–** Factors **"K"** and **"P"** both at first level.

$\boldsymbol{k_0 p_0 (or) 1}$ **– "K"** at second level and **"P"** at first level.

$\boldsymbol{k_0 p_0 (or) 1}$ **– "K"** at first level and **"P"** at second level

$\boldsymbol{k_0 p_0 (or) 1}$ **–** Factors **"K"** and **"P"** both at second level.

The above 4 treatment combinations can be compared by laying out the experiment in 4×4 Latin square design and ANOVA can be carried out with 4 – 1 degrees of freedom associated with the treatment effects. Our aim in the factorial experiment is to carry out separate tests for the main effects **K**, **P** and also the interaction **KP**, Splitting the **treatment S.S with D.f** into three orthogonal components, each with 1 D.f and each associated with the main effects **K** and **P** or the interaction **KP.**

NOTE:

1. $[\boldsymbol{K}] = [kp] - [p] + [k] - [1]$
2. $[\boldsymbol{P}] = [kp] + [p] - [k] - [1]$
3. $[\boldsymbol{KP}] = [kp] - [p] - [k] + [1]$
4. $sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{K} \to \boldsymbol{S_K} = \frac{[\boldsymbol{K}]^2}{\boldsymbol{4r}}$ **; r – no. of rows.**
5. $sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{P} \to \boldsymbol{S_P} = \frac{[\boldsymbol{P}]^2}{\boldsymbol{4r}}$
6. $sum\ of\ square\ due\ to\ interaction\ effect\ \ of\ \boldsymbol{KP} \to \boldsymbol{S_{KP}} = \frac{[\boldsymbol{KP}]^2}{\boldsymbol{4r}}$

   each with D.f = 1.

   Where, $[1], [K], [P], [KP]$ denote the total yields of r – units (plots) receiving treatments $1, k, p, kp$.
7. M = mean yield of the four treatment combinations then, $M = \frac{1}{4}(a+1)(b+1)$
8. **Table-II (ANOVA table):**

Source of variation	Degree of freedom	Sum of squares	Mean squares	F-Ratio
$\boldsymbol{k}$	1	S_K	$MS_K = \frac{S_K}{D.f}$	$F_C = \frac{MS_K}{MSE}$
$\boldsymbol{p}$	1	S_P	$MS_P = \frac{S_P}{D.f}$	$F_R = \frac{MS_P}{MSE}$
$\boldsymbol{kp}$	1	S_{KP}	$MS_{KP} = \frac{S_{KP}}{D.f}$	$F_{KP} = \frac{MS_{KP}}{MSE}$
Error	$N - c - r + 1$	SSE	MSE = $\frac{SSE}{D.f}$	*

Problems:

1. **An experiment was planned to study the effect of sulphate of potash & super phosphate on the yield of potatoes. All the combinations of 2 levels of super phosphate (0 cent (p_0) & 0 cent (p_1) / acre) were studied of potash (0 cent (k_0) and 0 cent (k_1) / acre) were studied in a randomized block design with 4 replications for each.The (1 / 70) yields (l_b per plot = $\left(\frac{1}{70}\right)$ acre) obtained are given in the following table. Analyze the data & give your conclusions at 1% level**

Block	**Yields (l_b per plot)**			
I	**(1)**	k	p	kp
	23	**25**	**22**	**38**
II	p	**(1)**	k	kp
	40	**26**	**36**	**38**
III	**(1)**	k	kp	p
	29	**20**	**30**	**20**
IV	kp	k	p	**(1)**
	34	**31**	**24**	**28**

Solution:

We re-arrange the given data in a new table as given below for computations of the SS due to treatments and blocks.

Treatment Combination	**Total yields - Blocks**			
	I	**II**	**III**	**IV**
(1)	**23**	**26**	**29**	**28**
k	**25**	**36**	**20**	**31**
p	**22**	**40**	**20**	**24**
kp	**38**	**38**	**30**	**34**

Solution: N -Total number of items in the given data = 16.

c - Number of columns = 4; r - Number of rows = 4.

Null Hypothesis H_0: The data is homogenous w. r. t blocks and treatments.

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 29 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Treatment Combination	**Total yields - Blocks**				**Row total R_j**	$\frac{R_j^2}{n}$	$\sum_j X_{ij}^2$
	I	**II**	**III**	**IV**			
(1)	-6	-3	0	-1	$-10[1]$	$\frac{(-10)^2}{4} = 25$	46
k	-4	7	-9	2	$-4\,[k]$	$\frac{(-4)^2}{4} = 4$	150
p	-7	11	-9	-5	$-10[p]$	$\frac{(-10)^2}{4} = 25$	276
kp	9	9	1	5	$24[kp]$	$\frac{(-24)^2}{4} = 144$	188
Column total C_i	-8	24	-17	1	**Grand total** G = 0	$\sum \frac{R_j^2}{n} = 198$	*
$\frac{C_i^2}{n}$	$\frac{(-8)^2}{4} = 16$	$\frac{(24)^2}{4} = 144$	$\frac{(-17)^2}{4} = 72.25$	$\frac{(1)^2}{4} = 0.25$	$\sum \frac{C_j^2}{n} = 232.5$	**C.F** = $\frac{G^2}{N} = \frac{(0)^2}{16} = 0$	*
$\sum_i X_{ij}^2$	182	260	163	55	*	*	$\sum\sum X_{ij}^2$ =660

Total sum of squares (T.S.S) = Sum of squares of all the items ($\sum\sum {X_{ij}}^2$) - Correction Factor(C.F)

$= 660 - 0 = 660$

Sum of squares between columns (SSC) $= \sum \frac{{C_i}^2}{n_i}$ - Correction Factor (C.F) = $232.5 - 0$ = 232.5

Sum of squares between rows (SSR) = $\sum \frac{{R_j}^2}{n_j}$ - Correction Factor (C.F) = $198 - 0 = 198$

SSE = Residual Sum of squares = TSS – (S.S.C + S.S.R) = 660 – (232.5+198) = 660 – 430.5 = 229.5

$[\boldsymbol{K}] = [kp] - [p] + [k] - [1] = 24 - (-10) + (-4) + 10 = 40$

$[\boldsymbol{P}] = [kp] + [p] - [k] - [1] = 24 - 10 - (-4) - (-10) = 28$

$[\boldsymbol{KP}] = [kp] - [p] - [k] + [1] = 24 - (-4) - (-10) + (-10) = 28$

$sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{K} \rightarrow \boldsymbol{S_K} = \frac{[\boldsymbol{K}]^2}{\boldsymbol{4r}} = \frac{[\boldsymbol{40}]^2}{\boldsymbol{4\times4}} = \frac{\boldsymbol{1600}}{\boldsymbol{16}} = \boldsymbol{100}$ **; r – no. of rows.**

$sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{P} \rightarrow \boldsymbol{S_P} = \frac{[\boldsymbol{P}]^2}{\boldsymbol{4r}} = \frac{[\boldsymbol{28}]^2}{\boldsymbol{4\times4}} = \frac{\boldsymbol{784}}{\boldsymbol{16}} = \boldsymbol{49}$

$sum\ of\ square\ due\ to\ interaction\ effect\ \ of\ \boldsymbol{KP} \rightarrow \boldsymbol{S_{KP}} = \frac{[\boldsymbol{KP}]^2}{\boldsymbol{4r}} = \frac{[\boldsymbol{28}]^2}{\boldsymbol{4\times4}} = \frac{\boldsymbol{784}}{\boldsymbol{16}} = \boldsymbol{49}$

each with D.f = 1.

Table-II (ANOVA table):

Source of variation	Degree of freedom	Sum of squares	Mean squares	F-Ratio
$\boldsymbol{k}$	1	$S_K = 100$	$\mathrm{MS}_K = \frac{S_K}{\mathrm{D.f}} = 100$	$\mathrm{F_K} = \frac{\mathrm{MS}_K}{\mathrm{MSE}} = 3.92$
$\boldsymbol{p}$	1	$S_P = 49$	$\mathrm{MS}_P = \frac{S_P}{\mathrm{D.f}} = 49$	$\mathrm{F_P} = \frac{\mathrm{MS}_P}{\mathrm{MSE}} = 1.92$
$\boldsymbol{kp}$	1	$S_{KP} = 49$	$\mathrm{MS}_{KP} = \frac{S_{KP}}{\mathrm{D.f}} = 49$	$\mathrm{F_{KP}} = \frac{\mathrm{MS}_{KP}}{\mathrm{MSE}} = 1.92$
Error	$N - c - r + 1 = 16 - 4 - 4 + 1 = 9$	SSE = 229.5	MSE = $\frac{\mathrm{SSE}}{\mathrm{D.f}} = \frac{229.5}{9} = 25.5$	*

Conclusion:

$\mathrm{F_{Tab}}$ for (1, 9) at 1% l.o.s is 10.56.

Since $\mathrm{F_{Calc}} < \mathrm{F_{Tab}}$ in all three cases H_0 is accepted. i.e., there are no significant effects present in the experiment due to different treatments.

2. **Find out the main effects and interaction in the following 2^2- Factorial experiment and write down the analysis of variance table at 1% l.o.s.**

Treatment Combination	Total yields - Blocks		
	I	**II**	**III**
(1)	**12**	**19**	**10**
$\boldsymbol{k}$	**15**	**20**	**16**
$\boldsymbol{p}$	**24**	**16**	**17**
$\boldsymbol{kp}$	**24**	**17**	**29**

Solution:

N -Total number of items in the given data = 12.

c - Number of columns = 3.

r - Number of rows = 4.

Null Hypothesis H_0: The data is homogenous w. r. t blocks and treatments (ie., There is no significant difference between blocks and treatments).

Shifting the origin does not affect the variance of a set of values, so we subtract (say) x = 20 from each value (to make the calculations simple).

Table – I (calculations for ANOVA):

Treatment Combination	Total yields - Blocks			Row total R_j	$\frac{R_j^{\,2}}{n}$	$\sum_j X_{ij}^{\,2}$
	I	II	III	***		
(1)	-8	-1	-10	$-19[1]$	$\frac{(-19)^2}{3} = 120.33$	165
$\boldsymbol{k}$	-5	0	-4	$-9\ [\boldsymbol{k}]$	$\frac{(-9)^2}{3} = 27$	41
$\boldsymbol{p}$	4	-4	-3	$-3[\boldsymbol{p}]$	$\frac{(-3)^2}{3} = 3$	41
$\boldsymbol{kp}$	4	-3	9	$10[\boldsymbol{kp}]$	$\frac{(10)^2}{3} = 33.33$	106
Column total C_i	-5	-18	-8	**Grand total** G = -21	$\sum \frac{R_j^{\,2}}{n} = 183.66$	*
$\frac{C_i^{\,2}}{n}$	$\frac{(-5)^2}{4} = 6.25$	$\frac{(-18)^2}{4} = 81$	$\frac{(-8)^2}{4} = 16$	$\sum \frac{C_j^{\,2}}{n} = 103.25$	**C.F** = $\frac{G^2}{N} = \frac{(-21)^2}{12} = \mathbf{36.75}$	*
$\sum_i X_{ij}^{\,2}$	121	26	206	*	*	$\sum\sum X_{ij}^{\,2}$ =353

Total sum of squares (T.S.S) = Sum of squares of all the items ($\sum\sum X_{ij}^{\,2}$) - Correction Factor(C.F)

$= 353 - 36.75 = 316.25$

Sum of squares between columns (SSC) $= \sum \frac{C_i^{\,2}}{n_i}$ - Correction Factor (C.F) = $103.25 - 36.75$

= 66.5

Sum of squares between rows (SSR) = $\sum \frac{R_j^{\,2}}{n_j}$ - Correction Factor (C.F) = $183.66 - 36.75$

$= 146.91$

SSE = Residual Sum of squares = TSS – (S.S.C + S.S.R) = 316.25 – (66.5+146.91)

= 316.25 – 213.41 = 102.84

$[\boldsymbol{K}] = [kp] - [p] + [k] - [1] = 10 - (-3) + (-9) - (-19) = 23$

$[\boldsymbol{P}] = [kp] + [p] - [k] - [1] = 10 + (-3) - (-9) - (-19) = 35$

$[\boldsymbol{KP}] = [kp] - [p] - [k] + [1] = 10 - (-9) - (-3) + (-19) = 3$

$sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{K} \rightarrow \boldsymbol{S_K} = \frac{[\boldsymbol{K}]^2}{\boldsymbol{4r}} = \frac{[\mathbf{23}]^2}{\mathbf{4 \times 3}} = \frac{\mathbf{529}}{\mathbf{12}} = \mathbf{44.08}$ **; r – no. of rows.**

$$sum\ of\ square\ due\ to\ main\ effect\ \boldsymbol{P} \rightarrow \boldsymbol{S_P} = \frac{[\boldsymbol{P}]^2}{\boldsymbol{4r}} = \frac{[\mathbf{35}]^2}{\mathbf{4 \times 3}} = \frac{\mathbf{1225}}{\mathbf{12}} = \mathbf{102.08}$$

$$sum\ of\ square\ due\ to\ interaction\ effect\ \ of\ \boldsymbol{KP} \rightarrow \boldsymbol{S_{KP}} = \frac{[\boldsymbol{KP}]^2}{\boldsymbol{4r}} = \frac{[\mathbf{3}]^2}{\mathbf{4 \times 3}} = \frac{\mathbf{9}}{\mathbf{12}} = \mathbf{0.75}$$

Table-II (ANOVA table):

Source of variation	Degree of freedom	Sum of squares	Mean squares	F-Ratio
$\boldsymbol{k}$	1	$S_K = 44.08$	$\mathrm{MS}_K = \frac{S_K}{\mathrm{D.f}} = 44.08$	$\mathrm{F_K} = \frac{\mathrm{MS}_K}{\mathrm{MSE}} = 2.57$
$\boldsymbol{p}$	1	$S_P = 102.08$	$\mathrm{MS}_P = \frac{S_P}{\mathrm{D.f}} = 102.08$	$\mathrm{F_P} = \frac{\mathrm{MS}_P}{\mathrm{MSE}} = 5.96$
$\boldsymbol{kp}$	1	$S_{KP} = 0.75$	$\mathrm{MS}_{KP} = \frac{S_{KP}}{\mathrm{D.f}} = 0.75$	$\mathrm{F_{KP}} = \frac{\mathrm{MS}_{KP}}{\mathrm{MSE}} = 22.85$

Error	$N - c - r + 1$ $= 12 - 3 - 4 + 1$ $= 6$	SSE = 102.33	MSE = $\frac{\text{SSE}}{\text{D.f}}$ $= \frac{102.33}{6} = 17.14$	*

Conclusion:

$\mathrm{F_{Tab}}$ for (1, 6) at 1% l.o.s is 13.75.

(i) $\mathrm{F_K} = 2.57$ Since $\mathrm{F_{Calc}} < \mathrm{F_{Tab}}$ H_0 is accepted. i.e., the mean effect of **K** is not significant.

(ii) $\mathrm{F_P} = 5.96$ Since $\mathrm{F_{Calc}} < \mathrm{F_{Tab}}$ H_0 is accepted. i.e., the mean effect of **P** is not significant.

(iii) $\mathrm{F_{KP}} = 22.85$ Since $\mathrm{F_{Calc}} > \mathrm{F_{Tab}}$ H_0 is rejected. i.e., the mean effect of interaction **KP** is significant at 1% l.o.s.