

Module-5

DESIGN OF EXPERIMENTS AND ANOVA

Principles of experimentation in design, Analysis of completely randomized design, randomized block design. The ANOVA Technique, Basic Principle of ANOVA, One-way ANOVA, Two-way ANOVA, Latin-square Design, and Analysis of Co-Variance.

DESIGN OF EXPERIMENTS:

Principles of experimentation in design:

There are three principles of experimental design.

(I) Principle of Randomization. (ii) Principle of replication and (iii) Principle of local control.

(i) Principle of Randomization: It involves the allocation of treatment to experimental units at random to avoid any bias in the experiment resulting from the influence of some extraneous unknown factor that may effect the experiment.

(ii) Principle of replication: The experiment should be repeated more than once. Thus, each treatment is applied in many experimental units instead of one.

(iii) Principle of local control: The control of all factors except the ones about which we are investigating.

Experimental Design (Design of experiment):

(i) Analysis of completely randomized design (CRD):

A completely randomized design is one where the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment.

(ii) Randomized block design: A randomized block design is an experimental design where the experimental units are in groups called blocks.

(iii) Latin Squares design: A Latin square is the arrangement of m treatments, each one repeated m times in such a way that each treatment appeared exactly one time in each row and each column in the design.

Examples of experimental design:

- (i) **Complete Randomization:** Suppose that an agricultural experiment consists of examining the yields per acre of four different varieties of wheat, where each variety is grown on four different plots of land. To design such an experiment, we could divide the land into $4 \times 4 = 16$ plots as shown in figure by square and assign each treatment indicated by A, B, C and D to four blocks chosen completely at random. The purpose of the randomization is to eliminate various sources of error, such as soil fertility.

Complete randomization:

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

- (ii) **Randomized blocks:** When as in the above example, it is necessary to have a complete set of treatments for each block, the treatments A, B, C and D are introduced in random order within each block: I, II, III and IV and for this reason blocks are referred to as randomized blocks. This type of design is used when it is desired to control one source of error or variability, namely, the difference in blocks.

Randomized blocks:

Blocks	Treatment			
I	C	B	A	D
II	A	B	D	C
III	B	C	D	A
IV	A	D	C	B

- (iii) **Latin squares:** For some purposes it is necessary to control two source of error or variability at same time, such as the difference in rows and the difference in columns. In the experiment of above example, for instance, error in different rows and columns could be due to changes in soil fertility in different parts of land. In such case it is desirable that each treatment occurs once in each row and once in each column, as shown in the below figure. The arrangement is called a Latin square. from the fact that Latin letters A, B, C and D are used.

Latin squares:

	Factor 1			
Factor 2	D	B	C	A
	B	D	A	C
	C	A	D	B
	A	C	B	D

The ANOVA Technique:

ANOVA is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors.

The systematic factors have a statistical influence on the given data set, while the random factors do not.

In other words, ANOVA is a statistical method that separates observed variance data into different components to use for additional tests.

Basic Principles of ANOVA:

Basic principle of ANOVA is to test for difference among the means of the populations by examining amount of variation within each of these samples, relative to the amount of variation between the samples.

One –way classification or one –Factor experiments (one way ANOVA):

Suppose we have m treatment (i.e., samples) and n measurements (i.e., observations) classified as below:

Treatment 1: $x_{11}, x_{12}, x_{13}, \dots, x_{1n} : \bar{x}_1$

Treatment 2: $x_{21}, x_{22}, x_{23}, \dots, x_{2n} : \bar{x}_2$

.....

.....

Treatment m: $x_{m1}, x_{m2}, x_{m3}, \dots, x_{mn} : \bar{x}_m$

Where \bar{x}_j is the mean of the measurements in the j^{th} row. We have $\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{jk}$,

Where $j = 1, 2, \dots, m$. The value \bar{x}_j are called group means, treatment means or row means.

The grand mean, or overall mean, is the mean of all the measurements in all the groups and is denoted by \bar{x} .

$$\therefore \bar{x} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n x_{jk}$$

Total variation:

We define the total variation, denoted by V, as the sum of the square of the deviation of each measurement from the grand mean \bar{x} .

$$V = \sum_{j,k} (x_{jk} - \bar{x})^2 \quad \text{i.e.,} \quad V = \sum_{j,k} (x_{jk} - \bar{x}_j)^2 + \sum_{j,k} (\bar{x}_j - \bar{x})^2 \quad \text{or}$$

$$V = \sum_{j,k} (x_{jk} - \bar{x}_j)^2 + n \sum_j (\bar{x}_j - \bar{x})^2 \quad \text{-----(1)}$$

Variation within treatment:

The first summation on the RHS of equation (1) is called the variation within treatments and it is denoted by V_w . $\therefore V_w = \sum_{j,k} (x_{jk} - \bar{x}_j)^2$ -----(2)

Variation between treatments:

The second summation on the RHS of equation (1) is called the variation between treatments and it is denoted by V_B . $\therefore V_B = \sum_{j,k} (\bar{x}_j - \bar{x})^2 = n \sum_j (\bar{x}_j - \bar{x})^2$ -----(3)

\therefore equation (1) becomes, Total variance $V = V_w + V_B$

Short cut methods for obtaining variations:

To minimize the labour of computing the above variations, we may use the following forms:

$$V = \sum_{j,k} x_{jk}^2 - \frac{T^2}{mn} \quad \text{-----(4).} \quad V_B = \frac{1}{n} \sum_j T_j^2 - \frac{T^2}{mn} \quad \text{-----(5).} \quad \text{And} \quad V_w = V - V_B \quad \text{-----(6).}$$

Where T is the total of all values x_{jk} and T_j is the total of all values in the j^{th} treatment.

$$\therefore T = \sum_{j,k} x_{jk} \quad \text{and} \quad T_j = \sum_k x_{jk} \quad \text{-----(7)}$$

NOTE:

In practice, it is convenient to subtract some fixed value from all the data in the table in order to simplify the calculations. This will not effect on the final results.

The F-Test for the null hypothesis of equal means:

To test for null hypothesis H_0 , that the sample means are equal. we define the F-distribution by

$$F = \frac{S_B^2}{S_w^2} \quad \text{Where} \quad S_B^2 = \frac{V_B}{m-1} \quad \text{and} \quad S_w^2 = \frac{V_w}{m(n-1)}$$

The statistic F has the F-distribution with $m - 1$ and $m(n - 1)$ degree of freedom.

If this statistic F is significantly large, we can conclude that there is a significant difference between the treatment means and can thus reject the hypothesis H_0 . Otherwise, we can accept H_0 or reserve judgement, pending further analysis.

Analysis of variance Table (ANOVA Table)

The calculations required for the above test are tabulated as below, which is known as ANOVA Table.

Variation	Degree of Freedom	Mean square	F
Between treatment $V_B = m \sum_j (\bar{x}_j - \bar{x})^2$	$m - 1$	$S_B^2 = \frac{V_B}{m - 1}$	$F = \frac{S_B^2}{S_w^2}$ with $m - 1$ and $m(n - 1)$ degrees of freedom.
Within treatment $V_w = V - V_B$ $V_w = \sum_{j,k} (x_{jk} - \bar{x}_j)^2$	$m(n - 1)$	$S_w^2 = \frac{V_w}{m(n - 1)}$	
Total $V = V_w + V_B$ $V = \sum_{j,k} (x_{jk} - \bar{x})^2$	$mn - 1$		

Problems:

- The following table shows the yields in bushels per acre of a certain variety of wheat grown in a particular type of soil treated with chemicals A, B or C. Find (i) mean yields for the different treatments, (ii) grand mean for all treatments, (iii) the total variation, (iv) The variation between treatments and the variation within treatments.

A	48	49	50	49
B	47	49	48	48
C	49	51	50	50

Write the ANOVA Table. Can we reject the null hypothesis of equal means at significance level of 0.05 and 0.01?

[Use $F_{0.05} = 4.26$ for df 2 and 9 and $F_{0.01} = 8.02$ for df 2 and 9]

Solution:

For our convenience we subtract some number, say 45, from all the data without affecting the value of the variations and we obtain the data table.

A	3	4	5	4
B	2	4	3	3
C	4	6	5	5

Here $m = 3$ and $n = 4$.

(i) Using $\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{jk}$, The treatment (row) means are $\bar{x}_1 = \frac{1}{4}(3 + 4 + 5 + 4) = 4$,

$$\bar{x}_2 = \frac{1}{4}(2 + 4 + 3 + 3) = 3, \quad \bar{x}_3 = \frac{1}{4}(4 + 6 + 5 + 5) = 5. \text{ Add 45 to each of these values.}$$

\therefore The means yields for different treatments are 49, 48, and 50 bushels per acre for A, B, and C respectively.

(ii) Using $\bar{x} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n x_{jk}$, The grand mean for all treatments is

$$\bar{x} = \frac{1}{12}(3 + 4 + 5 + 4 + 2 + 4 + 3 + 3 + 4 + 6 + 5 + 5) = 4.$$

\therefore The grand mean for the original set of data is $4 + 45 = 49$ bushels per acre.

(iii) The total variation is $V = \sum_{j,k} (x_{jk} - \bar{x})^2$

$$\therefore V = (3 - 4)^2 + (4 - 4)^2 + (5 - 4)^2 + (4 - 4)^2 + (2 - 4)^2 + (4 - 4)^2 + (3 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (6 - 4)^2 + (5 - 4)^2 + (5 - 4)^2 = 14.$$

(iv) The variation between treatment is

$$V_B = n \sum_j (\bar{x}_j - \bar{x})^2 = 4[(4 - 4)^2 + (3 - 4)^2 + (5 - 4)^2] = 8.$$

(v) The variation within treatment is

$$V_w = V - V_B = 14 - 8 = 6. \text{ Or}$$

$$V_w = \sum_{j,k} (x_{jk} - \bar{x}_j)^2 = (3 - 4)^2 + (4 - 4)^2 + (5 - 4)^2 + (4 - 4)^2 + (2 - 3)^2 + (4 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (4 - 5)^2 + (6 - 5)^2 + (5 - 5)^2 + (5 - 5)^2 = 6$$

ANOVA Table:

Variation	Degree of freedom	Mean square	F
Between treatment $V_B = 8$	$m - 1 = 2$	$S_B^2 = \frac{V_B}{m - 1} = \frac{8}{2} = 4$	$F = \frac{S_B^2}{S_w^2} = \frac{4}{2/3} = 6$ With 2 and 9 degrees of freedom.
Within treatment $V_w = 6$	$m(n - 1) = (3)(3) = 9$	$S_w^2 = \frac{V_w}{m(n - 1)} = \frac{6}{9} = \frac{2}{3}$	
Total $V = 14$	$mn - 1 = (3)(4) - 1 = 11$		

From Table, we have $F_{0.05} = 4.26$ for degree of freedom $\gamma_1 = 2$, $\gamma_2 = 9$ and $F_{0.01} = 8.02$ for degree of freedom $\gamma_1 = 2$, $\gamma_2 = 9$.

$$\therefore F > F_{0.05} \text{ and } F < F_{0.01}$$

\therefore we can reject the null hypothesis of equal means at 0.05 level and we cannot reject null hypothesis of equal means at 0.01 level.

2. A company wishes to purchase one of five different machines: A, B, C, D, and E. In an experiment designed to test whether there is a difference in the machine's performance, each of five experienced operators works on each of the machines for equal times. The following table shows the numbers of units produced by the machine. Test the hypothesis that there is no difference between the machines at significance levels of 0.05 and 0.01. [use $F_{0.05} = 2.87$ and $F_{0.01} = 4.43$ for degrees of freedom $\gamma_1 = 4$ and $\gamma_2 = 20$]

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Solution:

This solution is obtained by using short cut method alternatively.

For convenience we subtract by a suitable number, say 60, from all the data to obtain the following table:

Here $m = 5$ and $n = 5$.

						T_j	T_j^2
A	8	12	17	-18	-7	12	144
B	12	-7	3	-7	-12	-11	121
C	0	22	4	15	12	53	2809
D	-12	1	-3	4	-10	-20	400
E	4	5	10	8	-7	20	400
$\sum x_{jk}^2 = 2658$						54	$\sum_j T_j^2 = 3874$

$$\sum x_{jk}^2 = 64 + 144 + 289 + 324 + 49 + 144 + 49 + 9 + 49 + 144 + 0 + 484 + 16 + 22 + 144 + 144 + 1 + 9 + 16 + 100 + 16 + 25 + 100 + 64 + 49 = 2658.$$

$$\therefore V = \sum_{j,k} x_{j,k}^2 - \frac{T^2}{mn} = 2658 - \frac{(54)^2}{(5)(5)} = 2658 - 116.64 = 2541.3$$

$$V_B = \frac{1}{n} \sum_j T_j^2 - \frac{T^2}{mn} = \frac{1}{5} (3874) - \frac{(54)^2}{(5)(5)} = 774.8 - 116.64 = 658.16$$

$$V_w = V - V_B = 2541.36 - 658.16 = 1883.2$$

ANOVA Table:

Variation	Degree of freedom	Mean square	F
Between treatment $V_B = 658.2$	$m - 1 = 4$	$S_B^2 = \frac{V_B}{m-1} = \frac{658.2}{4} = 164.5$	$F = \frac{S_B^2}{S_w^2} = \frac{164.5}{94.16}$ $= 1.75$ With 4 and 20 degrees of freedom.
Within treatment $V_w = 1883.2$	$m(n - 1) = 5(4) = 20$	$S_w^2 = \frac{V_w}{m(n-1)} = \frac{1883.2}{20} = 94.16$	
Total $V = 2541.36$	$mn - 1 = (5)(5) - 1 = 24$		

From table, we have $F_{0.05} = 2.87$ and $F_{0.01} = 4.43$ for degrees of freedom $\gamma_1 = 4$, $\gamma_2 = 20$.

$\therefore F < F_{0.05}$ and $F < F_{0.01}$

\therefore Hence, we cannot reject the null hypothesis of equal means at 0.05 level and 0.01 level.

3. Set up an analysis of variance (ANOVA) table for the following per acre production data for three varieties of wheat, each grown on 4 plots and state if the variety difference are significant at 5% level. [Use $F_{0.05,2,9} = 4.26$]

Per acre production data			
Plot of land	Variety of wheat		
	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Solution:

Table for calculations:

	1	2	3	4	T_j	T_j^2
A	6	7	3	8	24	576
B	5	5	3	7	20	400
C	5	4	3	4	16	256
	$\sum x_{jk}^2 = 332$				$T=60$	$\sum_j T_j^2 = 1232$

$$\sum x_{jk}^2 = 6^2 + 7^2 + 3^2 + 8^2 + 5^2 + 5^2 + 3^2 + 7^2 + 5^2 + 4^2 + 3^2 + 4^2 = 332.$$

Here $m = 3$ and $n = 4$

$$\therefore V = \sum_{j,k} x_{j,k}^2 - \frac{T^2}{mn} = 332 - \frac{(60)^2}{(3)(4)} = 332 - 300 = 32.$$

$$V_B = \frac{1}{n} \sum_j T_j^2 - \frac{T^2}{mn} = \frac{1}{4} (1232) - \frac{(60)^2}{(3)(4)} = 308 - 300 = 8.$$

$$V_w = V - V_B = 32 - 8 = 24$$

ANOVA Table:

Variation	Degree of freedom	Mean square	F
Between treatment $V_B = 8$	$m - 1 = 2$	$S_B^2 = \frac{V_B}{m - 1} = \frac{8}{2} = 4$	$F = \frac{S_B^2}{S_w^2} = \frac{4}{2.67}$ $= 1.4981$ With 2 and 9 degrees of freedom.
Within treatment $V_w = 24$	$m(n-1) = 3(3) = 9$	$S_w^2 = \frac{V_w}{m(n-1)} = \frac{24}{9}$ $= 2.67$	
Total $V = 32$	$mn - 1 = (3)(4) - 1 = 11$		

From table, we have $F_{0.05} = 4.26$ for degree of freedom $\gamma_1 = 2$, $\gamma_2 = 9$.

$$\therefore F < F_{0.05}$$

\therefore Hence, we cannot reject the null hypothesis of equal means at 0.05 level.

Hence the difference in wheat output due to varieties is significant at 5% level.

4. A teacher wishes to test three different teaching method: I, II and III. To do this, three groups five students each are chosen at random and each group is taught by a different method. The same examination is given to all the students, and the grades in the table are obtained. Determine whether there is a difference between the teaching methods at significance level of 0.05 and 0.01. [Use $F_{0.05}(2, 12) = 3.89$ and $F_{0.01}(2, 12) = 6.93$]

Method I	75	62	71	58	73
Method II	81	85	68	92	90
Method III	73	79	60	75	81

Solution:

For convenience we subtract by a suitable number 80, from all the data to obtain the following table. Here $m = 3$ and $n = 5$.

Table for calculations:

						T_j	T_j^2
I	-5	-18	-9	-22	-7	-61	3721
II	1	5	-12	12	10	16	256
III	-7	-1	-20	-5	1	-32	1024
$\sum x_{jk}^2 = 1853$						$T = -77$	$\sum_j T_j^2 = 5001$

$$\sum x_{jk}^2 = 25 + 324 + 81 + 484 + 49 + 1 + 25 + 144 + 144 + 100 + 49 + 1 + 400 + 25 + 1 = 1853.$$

$$\therefore V = \sum_{j,k} x_{j,k}^2 - \frac{T^2}{mn} = 1853 - 395.27 = 1457.73$$

$$V_B = \frac{1}{n} \sum_j T_j^2 - \frac{T^2}{mn} = \frac{1}{5}(5001) - 395.27 = 604.93$$

$$V_w = V - V_B = 1457.73 - 604.93 = 852.8$$

ANOVA Table:

Variation	Degree of freedom	Mean square	F
Between treatment $V_B = 604.93$	$m - 1 = 2$	$S_B^2 = \frac{V_B}{m - 1} = \frac{604.93}{2} = 302.465$	$F = \frac{S_B^2}{S_w^2} = \frac{302.465}{71.067} = 4.2561$ With 2 and 12 degrees of freedom.
Within treatment $V_w = 852.8$	$m(n - 1) = 3(4) = 12$	$S_w^2 = \frac{V_w}{m(n - 1)} = \frac{852.8}{3(4)} = 71.067$	
Total $V = 1457.73$	$mn - 1 = (3)(5) - 1 = 14$		

From table, we have $F_{0.05}(2, 12) = 3.89$ and $F_{0.01}(2, 12) = 6.93$.

$$\therefore F > F_{0.05} \text{ and } F < F_{0.01}.$$

\therefore Null hypothesis is rejected at 0.05 level and accepted at 0.01 level.

\therefore Hence there is a difference in the three teaching methods at the 0.05 level, but not at the 0.01 level.

HOME WORK:

1. Three different kinds of food are tested on three groups of rats for 6 weeks. The objective is to check the difference in mean weight (in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to the following data: [Use $F_{0.05}(2, 15) = 3.68$]

Food 1	8	12	19	8	6	11
Food 2	4	5	4	6	9	7
Food 3	11	8	7	13	7	9

2. Three types of fertilizers are used on three groups of plants for 6 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply one-way ANOVA using a 0.05 significance level to the following data: [Use $F_{0.05}(2, 15) = 3.68$]

Fertilizer-1	6	8	4	5	3	4
Fertilizer-2	8	12	9	11	6	8
Fertilizer-3	13	9	11	8	7	12

Two-way classification or Two-factor experiments or Two-way ANOVA:

suppose we have m treatments and n blocks classified as below:

	Block			
	1	2	...	n
Treatment 1:	x_{11}	x_{12}	...	x_{1n}
Treatment 2:	x_{21}	x_{22}	...	x_{2n}
.		.		
.		.		
Treatment m:	x_{m1}	x_{m2}	...	x_{mn}
	\bar{x}_1	\bar{x}_2	...	\bar{x}_m

The mean of the entries in the j^{th} row is denoted by \bar{x}_j , where $j = 1, 2, \dots, m$ and the mean of the entries in the k^{th} column is denoted by \bar{x}_k , where $k = 1, 2, 3, \dots, n$. The overall, or grand mean is denoted by \bar{x} . In symbols,

$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{jk}, \quad \bar{x}_k = \frac{1}{m} \sum_{j=1}^m x_{jk}, \quad \bar{x} = \frac{1}{mn} \sum_{j,k} x_{j,k}$$

Total variation:

We define the total variation as $V = \sum_{j,k} (x_{jk} - \bar{x})^2$

Variation due to error or chance:

The variation due to error or chance is defined by $V_E = \sum_{j,k} (x_{jk} - \bar{x}_j - \bar{x}_k + \bar{x})^2$

V_E is also known as the residual variation or random variation.

Variation between treatments (row):

The variation between treatments is defined by $V_R = n \sum_{j=1}^m (\bar{x}_j - \bar{x})^2$

Variation between blocks(column):

The variation between blocks is defined by $V_C = m \sum_{k=1}^n (\bar{x}_k - \bar{x})^2$

∴ The total variation is $V = V_E + V_R + V_C$

Short cut methods for obtaining variations:

We obtain the variation by using shortcut methods as follows:

$$V = \sum_{j,k} x_{j,k}^2 - \frac{T^2}{mn}, \quad V_R = \frac{1}{n} \sum_{j=1}^m T_j^2 - \frac{T^2}{mn}, \quad V_C = \frac{1}{m} \sum_{k=1}^n T_k^2 - \frac{T^2}{mn}, \quad \text{and} \quad V_E = V - V_R - V_C$$

Where T_j is the total of entries in the j^{th} row i.e., $T_j = \sum_k x_{jk}$, T_k is the total of entries in the k^{th} column, i.e., $T_k = \sum_j x_{jk}$, and T is the total of all entries ie. $T = \sum_{j,k} x_{jk}$.

ANOVA Table:

Variation	Degrees of freedom	Mean square	F
Between treatment $V_R = n \sum_j (\bar{x}_j - \bar{x})^2$	$m - 1$	$S_R^2 = \frac{V_R}{m - 1}$	$F_1 = \frac{S_R^2}{S_E^2}$ with $(m - 1)$ and $(m - 1)(n - 1)$ degrees of freedom.
Within Blocks $V_C = m \sum_{k=1}^n (\bar{x}_k - \bar{x})^2$	$(n - 1)$	$S_C^2 = \frac{V_C}{(n - 1)}$	$F_2 = \frac{S_C^2}{S_E^2}$ with $(n - 1)$ and $(m - 1)(n - 1)$ degrees of freedom.
Residual or random $V_E = V - V_R - V_C$	$(m - 1)(n - 1)$	$S_E^2 = \frac{V_E}{(m - 1)(n - 1)}$	
Total $V = V_E + V_R + V_C$ $= \sum_{j,k} (x_{jk} - \bar{x})^2$	$mn - 1$		

Problems:

1. The following data shows the yields per acre of four different plant crops grown on lots treated with three different types of fertilizers. Apply two-way ANOVA to determine at the 0.05 significance level whether there is a difference in yield per acre (i) due to the fertilizers and (ii) due to the crops. [use $F_{0.05}(2, 6) = 5.14$ and $F_{0.05}(3, 6) = 4.76$]

	Crop I	Crop II	Crop III	Crop IV
Fertilizer A	4.5	6.4	7.2	6.7
Fertilizer B	8.8	7.8	9.6	7.0
Fertilizer C	5.9	6.8	5.7	5.2

Solution:

Table of calculation:

	Crop I	Crop II	Crop III	Crop IV	Row Total (T_j)	Row mean
Fertilizer A	4.5	6.4	7.2	6.7	24.8	$\bar{x}_1 = 6.2$
Fertilizer B	8.8	7.8	9.6	7.0	33.2	$\bar{x}_2 = 8.3$
Fertilizer C	5.9	6.8	5.7	5.2	23.6	$\bar{x}_3 = 5.9$
Column total (T_k)	19.2	21.0	22.5	18.9	Grand total $\sum_{j,k} x_{i,j} = 81.6$	
Column mean	$\bar{x}_1 = 6.4$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 7.5$	$\bar{x}_4 = 6.3$	Grand mean $\bar{x} = 6.8$	

Here $m = 3$, $n = 4$, $mn = 12$. Row means: $\bar{x}_1 = \frac{24.8}{4} = 6.2$, $\bar{x}_2 = 8.3$, $\bar{x}_3 = 5.9$

Column mean: $\bar{x}_1 = \frac{19.2}{3} = 6.4$, $\bar{x}_2 = 7.0$, $\bar{x}_3 = 7.5$, $\bar{x}_4 = 6.3$

Grand mean: $\bar{x} = \frac{1}{mn} \sum_{j,k} x_{i,j} = \frac{81.6}{12} = 6.8$

$$V_R = n \sum_j (\bar{x}_j - \bar{x})^2 = 4[(6.2 - 6.8)^2 + (8.3 - 6.8)^2 + (5.9 - 6.8)^2] = 13.68$$

$$V_C = m \sum_{k=1}^n (\bar{x}_k - \bar{x})^2 = 4[(6.4 - 6.8)^2 + (7.0 - 6.8)^2 + (7.5 - 6.8)^2 + (6.3 - 6.8)^2] = 2.82$$

Total variance V is

$$\begin{aligned} V &= \sum_{j,k} (x_{jk} - \bar{x})^2 = (4.5 - 6.8)^2 + (6.4 - 6.8)^2 + (7.2 - 6.8)^2 + (6.7 - 6.8)^2 \\ &\quad + (8.8 - 6.8)^2 + (7.8 - 6.8)^2 + (9.6 - 6.8)^2 + (7.0 - 6.8)^2 + (5.9 - 6.8)^2 \\ &\quad + (6.8 - 6.8)^2 + (5.7 - 6.8)^2 + (5.2 - 6.8)^2 = 23.08 \end{aligned}$$

$$\therefore V_E = V - V_R - V_C = 23.08 - 13.68 - 2.82 = 6.58$$

Short cut method (Alternative method):

Table of calculation:

	Crop I	Crop II	Crop III	Crop IV	Row Total (T_j)	T_j^2
Fertilizer A	4.5	6.4	7.2	6.7	24.8	615.04
Fertilizer B	8.8	7.8	9.6	7.0	33.2	1102.24
Fertilizer C	5.9	6.8	5.7	5.2	23.6	556.96
Column total (T_k)	19.2	21.0	22.5	18.9	T= 81.6	$\sum T_j^2 = 2274.24$
T_k^2	368.64	441.00	506.25	357.21	$\sum T_k^2 = 1673.1$	

From above table, we have,

$$\sum_{j,k} x_{j,k}^2 = (4.5)^2 + (6.4)^2 + (7.2)^2 + (6.7)^2 + (8.8)^2 + (7.8)^2 + (9.6)^2 + (7.0)^2 + (5.9)^2 + (6.8)^2 + (5.7)^2 + (5.2)^2 = 577.96$$

$$\therefore V = \sum_{j,k} x_{j,k}^2 - \frac{T^2}{mn} = 577.96 - \frac{(81.6)^2}{(3)(4)} = 23.08$$

$$V_R = \frac{1}{n} \sum_{j=1}^m T_j^2 - \frac{T^2}{mn} = \frac{1}{4} (2274.24) - \frac{(81.6)^2}{(3)(4)} = 13.68$$

$$V_C = \frac{1}{m} \sum_{k=1}^n T_k^2 - \frac{T^2}{mn} = \frac{1}{3} (1673.10) - \frac{(81.6)^2}{(3)(4)} = 2.82$$

$$\therefore V_E = V - V_R - V_C = 23.08 - 13.68 - 2.82 = 6.58$$

ANOVA Table:

Variation	Degrees of freedom	Mean square	F
Between treatments $V_R = 13.68$	$m - 1 = 3 - 1 = 2$	$S_R^2 = \frac{V_R}{m - 1} = \frac{13.68}{2} = 6.84$	$F_1 = \frac{S_R^2}{S_E^2} = \frac{6.84}{1.097} = 6.24$ with 2 and 6 degrees of freedom.
Between Blocks $V_C = 2.82$	$m - 1 = 4 - 1 = 3$	$S_C^2 = \frac{V_C}{(n - 1)} = \frac{2.82}{3} = 0.94$	$F_2 = \frac{S_C^2}{S_E^2} = \frac{0.94}{1.097} = 0.86$ with 3 and 6 degrees of freedom
Residual or random $V_E = 6.58$	$(m - 1)(n - 1) = 6$	$S_E^2 = \frac{V_E}{(m-1)(n-1)} = \frac{6.58}{6} = 1.097$	
Total $V = 23.08$	$mn - 1 = 12 - 1 = 11$		

∴ From table, $F_{0.05}(2,6) = 5.14$ ∴ $F_1 > F_{0.05}$.

∴ we rejected the hypothesis that the row means are equal and conclude that at 0.05 level there is a significant difference in yield due to the fertilizer.

Also, from table $F_{0.05}(3,6) = 4.76$ ∴ $F_2 < F_{0.05}$. ($F_2 < 1$)

∴ we conclude that there is no significant difference in yield due to the crops.

2. Set up an analysis of variance table for the following two-way ANOVA design results.

Per acre production data			
Varieties of fertilizers	Variety of seeds		
	A	B	C
W	6	5	5
X	7	5	4
Y	3	3	3
Z	8	7	4

Also state whether variety difference are significant at 5% level.

[Use $F_{0.05,3,6} = 4.76$ and $F_{0.05,2,6} = 5.14$]

Solution:

Here $m = 4$, $n = 3$

Table of calculations:

	A	B	C	T_j	T_j^2
W	6	5	5	16	256
X	7	5	4	16	256
Y	3	3	3	9	81
Z	8	7	4	19	361
T_k	24	20	16	$T = 60$	$\sum T_j^2 = 954$
T_k^2	576	400	256	$\sum T_k^2 = 1232$	

$$\sum_{j,k} x_{j,k}^2 = 36 + 25 + 25 + 49 + 25 + 16 + 9 + 9 + 9 + 64 + 49 + 16 = 332.$$

$$\therefore V = \sum x_{jk}^2 - \frac{T^2}{mn} = 332 - \frac{(60)^2}{(4)(3)} = 332 - 300 = 32$$

$$V_R = \frac{1}{n} \sum T_j^2 - \frac{T^2}{mn} = \frac{1}{3}(954) - 300 = 318 - 300 = 18$$

$$V_C = \frac{1}{m} \sum T_k^2 - \frac{T^2}{mn} = \frac{1}{4}(1232) - 300 = 308 - 300 = 8$$

$$\therefore V_E = V - V_R - V_C = 32 - 18 - 8 = 6.$$

ANOVA Table:

Variations	Degree of freedom	Mean square	F
Between varieties of Fertilizers $V_R = 18$	$m - 1 = 4 - 1 = 3$	$S_R^2 = \frac{V_R}{m - 1} = \frac{18}{3} = 6$	$F(3,6) = \frac{S_R^2}{S_E^2} = \frac{6}{1} = 6$
Between varieties of Seeds $V_C = 8$	$n - 1 = 3 - 1 = 2$	$S_C^2 = \frac{V_C}{n - 1} = \frac{8}{2} = 4$	$F(2,6) = \frac{S_C^2}{S_E^2} = \frac{4}{1} = 4$
Residual or error $V_E = 6$	$(m - 1)(n - 1) = 6$	$S_E^2 = \frac{V_E}{(m - 1)(n - 1)} = \frac{6}{6} = 1$	
Total $V = 18 + 8 + 6 = 32$	$mn - 1 = 12 - 1 = 11$		

From Table $F_{0.05}(3, 6) = 4.76$ and $F_{0.05}(2, 6) = 5.14$

$$\therefore F(3, 6) > F_{0.05}(3, 6).$$

\therefore Variety differences concerning fertilizers are significant at 5% level.

And $F(2, 6) < F_{0.05}(2, 6) = 5.14$.

\therefore Variety differences concerning fertilizers are not significant at 5% level.

3. Articles manufactured by a company are produced by 3 operators A, B, and C using 3 different machines X, Y, and Z. The manufacturer wishes to determine whether there is a difference between the operators and between the machines. An experiment is performed to determine the number of articles per day produced by each operator using each machine and the results are tabulated as below. Test if the mean production is equal with respect to (i) machines (ii) operators. Given that $F_{0.05} = 6.94$ for (2, 4) df.

	A	B	C
Machine X	23	27	24
Machine Y	34	30	28
Machine Z	28	25	27

Solution:

Here $m = 3$, $n = 3$. For convenience we subtract all the data by a suitable number 25.

Table of calculations:

	A	B	C	T_j	T_j^2
X	-2	2	-1	-1	1
Y	9	5	3	17	289
Z	3	0	2	5	25
T_k	10	7	4	$T = 21$	$\sum T_j^2 = 315$
T_k^2	100	49	16	$\sum T_k^2 = 165$	

$$\sum x_{jk}^2 = 4 + 4 + 1 + 81 + 25 + 9 + 9 + 0 + 4 = 137$$

$$\therefore V = \sum x_{jk}^2 - \frac{T^2}{mn} = 137 - \frac{(21)^2}{(3)(3)} = 137 - 49 = 88$$

$$V_R = \frac{1}{n} \sum T_j^2 - \frac{T^2}{mn} = \frac{1}{3}(315) - 49 = 56$$

$$V_C = \frac{1}{m} \sum T_k^2 - \frac{T^2}{mn} = \frac{1}{3}(165) - 49 = 6$$

$$\therefore V_E = V - V_R - V_C = 88 - 56 - 6 = 26$$

ANOVA Table:

Variations	Degree of freedom	Mean square	F
Between the machines $V_R = 56$	$m - 1 = 3 - 1 = 2$	$S_R^2 = \frac{V_R}{m - 1} = \frac{56}{2} = 28$	$F_1(2,4) = \frac{S_R^2}{S_E^2} = \frac{28}{6.5} = 4.31$
Between the operators $V_C = 6$	$n - 1 = 3 - 1 = 2$	$S_C^2 = \frac{V_C}{n - 1} = \frac{6}{2} = 3$	$F_2(2,4) = \frac{S_C^2}{S_E^2} = \frac{3}{6.5} = 0.46$
Residual or error $V_E = 26$	$(m - 1)(n - 1) = 4$	$S_E^2 = \frac{V_E}{(m - 1)(n - 1)} = \frac{26}{4} = 6.5$	
Total $V = 56 + 6 + 26 = 88$	$mn - 1 = 9 - 1 = 8$		

From table $F_{0.05}(2,4) = 6.94$. $\therefore F_1(2,4) < F_{0.05}(2,4)$ and $F_2(2,4) < F_{0.05}(2,4)$

\therefore The hypothesis is accepted. Hence there is no difference in the operators or machines at the 0.05 level.

HOME WORK:

4. The following table gives the monthly sales (in thousand rupees) of a certain firm in three states by its four salesmen:

States	Salesmen			
	A	B	C	D
X	5	4	4	7
Y	7	8	5	4
Z	9	6	6	7

Set up an analysis of variance table for the above data. Calculate F-coefficients and state whether the difference between sales affected by the four salesmen and difference between sales affected in three states are significant at 5% level of significance.

[Use $F_{0.05,2,6} = 5.14$ and $F_{0.05,3,6} = 4.76$]

Latin Squares:

Problems:

1. A farmer wishes to test the effects of four different fertilizers A, B, C, and D on the yield of wheat. In order to eliminate sources of error due to variability in soil fertility, he uses the fertilizers in a Latin-square arrangement, as shown in the following table, where the numbers indicate the yields in bushels per unit area. Perform an analysis of variance to determine whether there is a difference between the fertilizers at significance levels of (i) 0.05 and (ii) 0.01. [Use $F_{0.05,3,6} = 4.76$ and $F_{0.01,3,6} = 9.78$]

A 18	C 21	D 25	B 11
D 22	B 12	A 15	C 19
B 15	A 20	C 23	D 24
C 22	D 21	B 10	A 17

Solution:

Here $m = 4$ and $n = 4$.

For convenience we subtract all the data by a suitable number 20.

Table of calculations:

						T_j	T_j^2			
	A	-2	C	1	D	5	B	-9	-5	25
	D	2	B	-8	A	-5	C	-1	-12	144
	B	-5	A	0	C	3	D	4	2	4
	C	2	D	1	B	-10	A	-3	-10	100
T_k	-3		-6		-7		-9		T = -25	$\sum T_j^2 = 273$
T_k^2	9		36		49		81		$\sum T_k^2 = 175$	

Let r denote the number of treatments. $\therefore r = 4$.

Treatments Table:

	A	B	C	D	
	-2	-5	2	2	
	0	-8	1	1	
	-5	-10	3	5	
	-3	-9	-1	4	
T_t	-10	-32	5	12	
T_t^2	100	1024	25	144	$\sum T_t^2 = 1293$

$$\sum x_{jk}^2 = 4 + 1 + 25 + 81 + 4 + 64 + 25 + 1 + 25 + 0 + 9 + 16 + 4 + 1 + 100 + 9 = 369.$$

The total variance is

$$V = \sum x_{jk}^2 - \frac{T^2}{mn} = 369 - 39.06 = 329.94$$

The variation between rows is

$$V_R = \frac{1}{n} \sum T_j^2 - \frac{T^2}{mn} = \frac{1}{4} (273) - \frac{(-25)^2}{(4)(4)} = 68.25 - 39.06 = 29.19$$

The variation between columns is

$$V_C = \frac{1}{m} \sum T_k^2 - \frac{T^2}{mn} = \frac{1}{4} (175) - \frac{(-25)^2}{(4)(4)} = 43.75 - 39.06 = 4.69$$

The variation between treatments is

$$V_B = \frac{1}{r} \sum T_t^2 - \frac{T^2}{mn} = \frac{1}{4} (1293) - 39.06 = 323.25 - 39.06 = 284.19$$

\therefore Residuals variation is

$$V_E = V - V_R - V_C - V_B = 329.94 - 29.19 - 4.69 - 284.19 = 11.87$$

ANOVA Table:

Variations	Degrees of freedom	Mean square	F
Between rows $V_R = 29.19$	$m - 1 = 4 - 1 = 3$	$S_R^2 = \frac{V_R}{m - 1} = \frac{29.19}{3} = 9.73$	$F_1(3,6) = \frac{S_R^2}{S_E^2} = \frac{9.73}{1.978} = 4.919$
Between columns $V_c = 4.69$	$n - 1 = 4 - 1 = 3$	$S_c^2 = \frac{V_c}{(n - 1)} = \frac{4.69}{3} = 1.563$	$F_2(3,6) = \frac{S_c^2}{S_E^2} = \frac{1.563}{1.978} = 0.79$
Between Treatment $V_B = 284.19$	$r - 1 = 4 - 1 = 3$	$S_B^2 = \frac{V_B}{(r - 1)} = \frac{284.19}{3} = 94.73$	$F_3(3,6) = \frac{S_B^2}{S_E^2} = \frac{94.73}{1.978} = 47.89$
Residuals $V_E = 11.87$	$(m-1)(m-2) = (3)(2) = 6$	$S_E^2 = \frac{V_E}{(m-1)(m-2)} = \frac{11.87}{6} = 1.978$	
Total $V=329.94$	$mn - 1 = 16 - 1 = 15$		

\therefore From table, $F_{0.05}(3,6) = 4.76$ and $F_{0.01}(3,6) = 9.78$

(i) $F_1(3,6) > F_{0.05}$. \therefore We can reject the hypothesis that there are equal row means. It follows that there is a difference in the fertility of the soil from one row to another at 0.05 level.

$F_2 < F_{0.05}$. \therefore There is no difference in the soil fertility in the column.

$F_3 > F_{0.05}$. \therefore There is a difference between fertilizers.

(ii) $F_1 < F_{0.01}$. $F_2 < F_{0.01}$. And $F_3 > F_{0.01}$. \therefore There is no difference in the fertility of the soil in the rows at 0.01 level, there is no difference in the soil fertility in the columns and there is a difference between fertilizer at the 0.01 level.

2. Analyze and interpret the following statistics concerning output of wheat per field obtained as a result of experiment conducted to test four varieties of wheat viz., A, B, C and D under a Latin- square design. [Use $F_{0.05,3,6} = 4.76$]

C 25	B 23	A 20	D 20
A 19	D 19	C 21	B 18
B 19	A 14	D 17	C 20
D 17	C 20	B 21	A 21

Solution:

Here $m = 4$, $n = 4$, $r = \text{number of varieties} = 4$

For convenience, we subtract all the data by a suitable number say 20.

Table of calculation:

					T_j	T_j^2
	C 5	B 3	A 0	D 0	8	64
	A -1	D -1	C 1	B -2	-3	9
	B -1	A -6	D -3	C 0	-10	100
	D -3	C 0	B 1	A -5	-7	49
T_k	0	-4	-1	-7	$T = -12$	$\sum T_j^2 = 222$
T_k^2	0	16	1	49	$\sum T_k^2 = 66$	

Varieties Table:

	A	B	C	D	
	-1	-1	5	-3	
	-6	3	0	-1	
	0	1	1	-3	
	-5	-2	0	0	
T_t	-12	1	6	-7	
T_t^2	144	1	36	49	$\sum T_t^2 = 230$

$$\sum x_{jk}^2 = 25 + 9 + 0 + 0 + 1 + 1 + 1 + 4 + 1 + 36 + 9 + 0 + 9 + 0 + 1 + 25 = 122.$$

The total variation is

$$V = \sum x_{jk}^2 - \frac{T^2}{mn} = 122 - \frac{(-12)^2}{(4)(4)} = 122 - 9 = 113.$$

The variation between rows is

$$V_R = \frac{1}{n} \sum T_j^2 - \frac{T^2}{mn} = \frac{1}{4} (222) - \frac{(-12)^2}{(4)(4)} = 55.5 - 9 = 46.5$$

The variation between columns is

$$V_C = \frac{1}{m} \sum T_k^2 - \frac{T^2}{mn} = \frac{1}{4} (66) - 9 = 16.5 - 9 = 7.5$$

The variation between varieties is

$$V_B = \frac{1}{r} \sum T_t^2 - \frac{T^2}{mn} = \frac{1}{4} (230) - 9 = 57.5 - 9 = 48.5$$

\therefore Residuals variation is

$$V_E = V - V_R - V_C - V_B = 113 - 46.5 - 7.5 - 48.5 = 10.5$$

ANOVA Table:

Variations	Degrees of freedom	Mean square	F
Between rows $V_R = 46.5$	$m - 1 = 4 - 1 = 3$	$S_R^2 = \frac{V_R}{m - 1} = \frac{46.5}{3} = 15.5$	$F_1(3,6) = \frac{S_R^2}{S_E^2} = \frac{15.5}{1.75} = 8.857$
Between columns $V_c = 7.5$	$n - 1 = 4 - 1 = 3$	$S_c^2 = \frac{V_c}{(n - 1)} = \frac{7.5}{3} = 2.5$	$F_2(3,6) = \frac{S_c^2}{S_E^2} = \frac{2.5}{1.75} = 1.429$
Between Treatment $V_B = 48.5$	$r - 1 = 4 - 1 = 3$	$S_B^2 = \frac{V_B}{(r - 1)} = \frac{48.5}{3} = 16.17$	$F_3(3,6) = \frac{S_B^2}{S_E^2} = \frac{16.17}{1.75} = 9.24$
Residuals $V_E = 10.5$	$(m - 1)(m - 2) = (3)(2) = 6$	$S_E^2 = \frac{V_E}{(m - 1)(m - 2)} = \frac{10.5}{6} = 1.75$	
Total $V=113$	$mn - 1 = 16 - 1 = 15$		

∴ From table, $F_{0.05}(3,6) = 4.76$

$F_1 > F_{0.05}$, $F_2 < F_{0.05}$, and $F_3 > F_{0.05}$.

∴ The variance between rows and variance between varieties are significant at 5% level of significance and the variance between columns a not significant at 5% level of significance.

3. An experiment is performed to test the effect on corn yield of four different fertilizer treatments A, B, C, and D and of soil variations in two perpendicular directions. The following Latin-square table is obtained, where the numbers indicate the corn yield per unit area. Test at the 0.01 significance levels of the hypothesis that there is no difference between (i) the fertilizers and (ii) the soil variations. [Use $F_{0.01,3,6} = 9.78$]

C 8	A 10	D 12	B 11
A 14	C 12	B 11	D 15
D 10	B 14	C 16	A 10
B 7	D 16	A 14	C 12

Solution:

Here $m = 4$, $n = 4$, $r = \text{number of varieties} = 4$

For convenience, we subtract all the data by a suitable number say 10.

Table of calculation:

					T_j	T_j^2
	C -2	A 0	D 2	B 1	1	1
	A 4	C 2	B 1	D 5	12	144
	D 0	B 4	C 6	A 0	10	100
	B -3	D 6	A 4	C 2	9	81
T_k	-1	12	13	8	$T = 32$	$\sum T_j^2 = 326$
T_k^2	1	144	169	64	$\sum T_k^2 = 378$	

Varieties Table:

	A	B	C	D	
	4	-3	-2	0	
	0	4	2	6	
	4	1	6	2	
	0	1	2	4	
T_t	8	3	8	13	
T_t^2	64	9	64	169	$\sum T_t^2 = 306$

$$\sum x_{jk}^2 = 4 + 0 + 4 + 1 + 16 + 4 + 1 + 25 + 0 + 16 + 36 + 0 + 9 + 36 + 16 + 4 = 172.$$

The total variation is

$$V = \sum x_{jk}^2 - \frac{T^2}{mn} = 172 - \frac{(32)^2}{(4)(4)} = 172 - 64 = 108.$$

The variation between rows is

$$V_R = \frac{1}{n} \sum T_j^2 - \frac{T^2}{mn} = \frac{1}{4} (326) - \frac{(32)^2}{(4)(4)} = 81.5 - 64 = 17.5$$

The variation between columns is

$$V_C = \frac{1}{m} \sum T_k^2 - \frac{T^2}{mn} = \frac{1}{4} (378) - \frac{(32)^2}{(4)(4)} = 94.5 - 64 = 30.5$$

The variation between varieties is

$$V_B = \frac{1}{r} \sum T_t^2 - \frac{T^2}{mn} = \frac{1}{4} (306) - \frac{(32)^2}{(4)(4)} = 76.5 - 64 = 12.5$$

\therefore Residuals variation is

$$V_E = V - V_R - V_C - V_B = 108 - 17.5 - 30.5 - 12.5 = 47.5$$

ANOVA Table:

Variations	Degrees of freedom	Mean square	F
Between rows $V_R = 17.5$	$m - 1 = 4 - 1 = 3$	$S_R^2 = \frac{V_R}{m - 1} = \frac{17.5}{3} = 5.83$	$F_1(3,6) = \frac{S_R^2}{S_E^2} = \frac{5.83}{7.92} = 0.74$
Between columns $V_C = 30.5$	$n - 1 = 4 - 1 = 3$	$S_C^2 = \frac{V_C}{(n - 1)} = \frac{30.5}{3} = 10.17$	$F_2(3,6) = \frac{S_C^2}{S_E^2} = \frac{10.17}{7.92} = 1.28$
Between Treatment $V_B = 12.5$	$r - 1 = 4 - 1 = 3$	$S_B^2 = \frac{V_B}{(r - 1)} = \frac{12.5}{3} = 4.17$	$F_3(3,6) = \frac{S_B^2}{S_E^2} = \frac{4.17}{7.92} = 0.53$
Residuals $V_E = 47.5$	$(m - 1)(m - 2) = (3)(2) = 6$	$S_E^2 = \frac{V_E}{(m - 1)(m - 2)} = \frac{47.5}{6} = 7.92$	
Total $V = 108$	$mn - 1 = 16 - 1 = 15$		

∴ From table, $F_{0.01}(3,6) = 9.78$

$F_1 < F_{0.01}$, $F_2 < F_{0.01}$, and $F_3 < F_{0.01}$.

∴ Hypothesis is accepted. i.e., There is no significance difference between (i) the fertilizers and (ii) the soil variations at 0.01 level.

HOME WORK:

4. Present your conclusions after doing analysis of variance to the following results of the Latin-square design experiment in respect of five fertilizers which were used on plots of different fertility. [Use $F_{0.05}(4, 12) = 3.26$]

A 16	B 10	C 11	D 09	E 09
E 10	C 09	A 14	B 12	D 11
B 15	D 08	E 08	C 10	A 18
D 12	E 06	B 13	A 13	C 12
C 13	A 11	D 10	E 07	B 14