

# Design of Experiments

SESSION  
**12**

## Structure

- 12.1 Introduction
- Objectives
- 12.2 Problem Description
- 12.3 Completely Randomised Design
- 12.4 Randomised Block Design
- 12.5 Latin Square Design

## 12.1 INTRODUCTION

In Lab Session 11, you have learnt how to apply one-way and two-way ANOVA (with one observation per cell and m observations per cell) using MS Excel 2007. In Block 3 of MST-005, you have learnt that in any field of study such as agriculture, life sciences, industry, etc., it is essential to plan an experiment, that is, decide the objective of the experiment and the type of data required. In order to make efficient use of time and energy, experiments should be planned and designed carefully. Once the design of an experiment is finalised, observations are made and the data so obtained are analysed with the help of analysis of variance (ANOVA).

In this lab session, you will learn how to analyse the data of completely randomised design (CRD), randomised block design (RBD) and Latin square design (LSD) using MS Excel 2007.

### Objectives

After performing the activities of this session, you should be able to:

- prepare the spreadsheet in MS Excel 2007;
- analyse the data of completely randomised design (CRD);
- analyse the data of randomised block design (RBD); and
- analyse the data of Latin square design (LSD).

### Prerequisite

- Lab Session 11 of MSTL-001 (Basic Statistics Lab).
- Block 3 of MST-005 (Statistical Techniques).

## 12.2 PROBLEM DESCRIPTION

In this lab session, we state three problems to illustrate the applications of different types of design of experiments:

1. An experiment is conducted using completely randomised design to see the effect of four types of diets A, B, C and D on the gain in weights of rabbits. In this experiment, these diets are replicated 5 times randomly. Observations taken on the gains in weights due to different diets are given in Table 1.

**Table 1: Gain in weight**

60(B)	37(A)	70(D)	82(B)	83(D)
64(C)	59(D)	50(A)	66(C)	67(B)
58(D)	58(B)	63(A)	72(C)	62(C)
52(A)	87(C)	75(D)	56(B)	40(A)

Assuming that the gains in weights due to each diet are normally distributed with approximately equal variances, analyse the data to determine whether these diets differ significantly with respect to the gains in weights of rabbits at 5% level of significance.

2. An analyst of a textile company is interested in comparing the effect of four different chemicals (C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub> and C<sub>4</sub>) on the strength (y) of six bolts of cloth. Since there might be variability from one bolt to another, the analyst decides to use the randomised block design. She randomly selects a strip of cloth from each bolt and cuts it into four pieces (samples). The pieces are randomly assigned to receive one of the four chemicals (treatments). The analyst measures the resulting strengths (y) for each piece (sample). The data are given in the following table:

**Table 2: Strength of the cloth**

Bolt (Block)					
72 (C <sub>3</sub> )	72 (C <sub>1</sub> )	70 (C <sub>2</sub> )	71 (C <sub>1</sub> )	68 (C <sub>2</sub> )	72 (C <sub>1</sub> )
75 (C <sub>1</sub> )	68(C <sub>4</sub> )	69 (C <sub>3</sub> )	70 (C <sub>2</sub> )	66 (C <sub>4</sub> )	69 (C <sub>3</sub> )
70 (C <sub>4</sub> )	72 (C <sub>2</sub> )	74 (C <sub>1</sub> )	68 (C <sub>4</sub> )	72 (C <sub>3</sub> )	67 (C <sub>4</sub> )
73 (C <sub>2</sub> )	70 (C <sub>3</sub> )	70 (C <sub>4</sub> )	72 (C <sub>3</sub> )	70 (C <sub>1</sub> )	70 (C <sub>2</sub> )

The effect of each chemical on the strength (y) of six bolts of cloth is normally distributed with approximately equal variances.

- i) Analyse the design at 1% level of significance.
- ii) Is the average effect of different chemicals on the strength of cloth significantly different? If the difference between the effect of chemicals is significant, do the pair-wise comparison between them.
- iii) Is the effect of the bolts on the strength of cloth the same for all bolts?
3. An agricultural scientist designed an experiment to determine the effect of four fertilisers (A, B, C and D) on the yield of a crop. The other recognised sources of variation were soil fertility level and water level. The scientist randomly selected a 4 × 4 Latin square and randomly arranged the results of the experiment for different fertilisers. The data are presented in the following table:

**Table 3: Yield of the crop**

Soil fertility level	Water level			
	I	II	III	IV
1	14.55(D)	9.45(B)	14.70(C)	2.85(A)
2	5.20(C)	12.10(A)	17.50(D)	9.55(B)
3	2.70(A)	15.40(D)	12.00(B)	16.50(C)
4	12.45(B)	20.85(C)	4.75(A)	14.45(D)

Test at 5% level of significance:

- i) whether the effect of the soil fertility levels on the yield of the crop is significantly different by assuming that the effect of each soil fertiliser is normally distributed with approximately equal variance.
- ii) whether the effect of the water levels on the yield of the crop is significantly different by assuming that the effect of each water level is normally distributed with approximately equal variance.
- iii) whether the effect of different fertilisers on the yield of the crop is significantly different by assuming that the effect of each fertiliser is normally distributed with approximately equal variance.

## 12.3 COMPLETELY RANDOMISED DESIGN

In Unit 9 of MST-005, you have learnt that the completely randomised design (CRD) is frequently used when the experimental units are homogeneous. It means that the units have no inherent variation which could influence the effect of the treatments. For example, if we wish to apply the CRD in feeding experiment of cows, the cows (experimental units) should be homogeneous in all respects, i.e., the cows should be of the same breed, same age, same location, etc. In CRD, the treatments are allocated on the experimental units randomly and any treatment can be replicated any number of times.

The analysis of CRD is done with the help of one-way ANOVA, which has been described in Unit 9. We briefly mention the main steps and formulae as follows:

**Step 1:** We first formulate the null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ). If there are  $k$  treatments (or a treatment has  $k$  levels) and  $\mu_i$  ( $i = 1, 2, \dots, k$ ) is the average effect of the  $i^{\text{th}}$  treatment, we can take the null and alternative hypotheses as follows:

$H_0$ : The average effect of all treatments is equal or the average effect of the treatments is not significantly different

$H_1$ : The average effect of all treatments is not equal or the average effect of the treatments is significantly different

Symbolically,

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \text{At least one } \mu_i \neq \mu_j \ (i \neq j = 1, 2, \dots, k)$$

**Step 2:** We arrange the data of the CRD in the format of the one-way ANOVA, i.e., we arrange the data treatment-wise.

**Step 3:** We analyse the CRD with the help of one-way ANOVA as explained in the screened box in Sec. 11.3 of Lab Session 11.

**Step 4:** If the null hypothesis for the treatments is rejected, we do pair-wise comparison as explained in Sec. 11.3 of Lab Session 11.

Treatments – The conditions or factors whose effects are measured or compared are called treatments.

Experimental units – A unit to which a treatment is applied is called an experimental unit.

Experimental material – A set/group of experimental units is called experimental material.

### Steps in Excel

In Problem 1, the experiment is conducted using CRD and we have to test whether the diets differ significantly with respect to the gains in weights of rabbits.

We first have to set up the null and alternative hypotheses. If  $\mu_A$ ,  $\mu_B$ ,  $\mu_C$  and  $\mu_D$  denote the average gains in weights of rabbits due to diets A, B, C and D, respectively, we can formulate the null and alternative hypotheses as follows:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D, \text{ i.e., the diets do not differ significantly}$$

$$H_1 : \text{The diets differ significantly (claim)}$$

The following steps are used to analyse the CRD with the help of one-way ANOVA using the **Data Analysis ToolPak** in MS Excel 2007:

**Step 1:** We enter the data (given in Table 1) diet-wise as per the format of the one-way ANOVA in an Excel 2007 spreadsheet. We start by entering the heading of the data in Row 1 in Excel sheet and the observations of diets A, B, C and D in Columns A, B, C and D, respectively, as shown in Fig. 12.1.

	A	B	C	D	E
1	A	B	C	D	
2	37	60	64	70	
3	50	82	66	83	
4	63	67	72	59	
5	52	58	62	58	
6	40	56	87	75	
7					

Fig. 12.1: Partial screenshot of the spreadsheet for the given data.

**Step 2:** We repeat Steps 2 and 3 under the heading ‘Steps in Excel’ in Sec.11.3 of Lab Session 11 and obtain the results shown in Fig. 12.2.

	A	B	C	D	E	F	G
7	Anova: Single Factor						
8							
9	SUMMARY						
10	Groups	Count	Sum	Average	Variance		
11	A	5	242	48.4	107.3		
12	B	5	323	64.6	111.8		
13	C	5	351	70.2	102.2		
14	D	5	345	69	113.5		
15							
16							
17	ANOVA						
18	Source of Variation	SS	df	MS	F	P-value	F crit
19	Between Groups	1517.75	3	505.9167	4.6542	0.0160	3.2389
20	Within Groups	1739.2	16	108.7			
21							
22	Total	3256.95	19				
23							

Fig. 12.2

The explanation for these results is the same as given in Step 3 under the heading ‘Steps in Excel’ in Sec. 11.3 of Lab Session 11.

#### Decision using the critical region approach

Note that the calculated F value ( $F_{cal}$ ) is 4.6542 given in Cell E19, which is greater than the critical value 3.2389 given in Cell G19. It means that  $F_{cal}$  lies in the rejection region. So we reject the null hypothesis. Since the alternative hypothesis is the claim, we do not reject the claim. Hence, we conclude that there is not enough evidence against the claim. So we may conclude that the diets differ significantly with respect to the gains in the weights of rabbits at 5% level of significance.

#### Decision using the p-value approach

Since the p-value ( $= 0.0160$ ) is less than  $\alpha$  ( $= 0.05$ ), we reject the null hypothesis.

**Step 3:** Since the null hypothesis is rejected, i.e., there is a significantly difference between the diets with respect to the gains in the weights of

rabbits, we do pair-wise comparison as explained in Step 4 under the heading ‘Steps in Excel’ in Sec.11.3 of Lab Session 11. The final output is shown in Fig. 12.3.

	A	B	C	D	E	F	G	H	I
24	Comparison table								
25	Pair of treatments	n <sub>i</sub>	n <sub>j</sub>	$\bar{T}_i$	$\bar{T}_j$	$ \bar{T}_i - \bar{T}_j $	t <sub>(16),0.025</sub>	CD	Inference
26	A, B	5	5	48.4	64.6	16.2	2.1199	13.9785	Significant
27	A, C	5	5	48.4	70.2	21.8	2.1199	13.9785	Significant
28	A, D	5	5	48.4	69	20.6	2.1199	13.9785	Significant
29	B, C	5	5	64.6	70.2	5.6	2.1199	13.9785	Insignificant
30	B, D	5	5	64.6	69	4.4	2.1199	13.9785	Insignificant
31	C, D	5	5	70.2	69	1.2	2.1199	13.9785	Insignificant

Fig. 12.3

From the comparison table shown in Fig. 12.3, we observe that diet A for the gains in weights of rabbits differs significantly from diets B, C and D. However, diet B is not significantly different from diets C and D. Also the diet C does not significantly differ from diet D.

## 12.4 RANDOMISED BLOCK DESIGN

In Unit 10 of MST-005, you have learnt that the completely randomised design (CRD) is used when the experimental units are homogeneous. But such an ideal condition (that the experimental units are fully homogeneous) is rare. If there is only one factor which is likely to influence the effect of the treatment, then for controlling the variability of the experimental units, similar units are grouped together. These groups are known as **blocks** and each block is homogeneous. For example, in feeding experiment of cows, if the cows are uniform with respect to age and breed but are from different locations, we cannot apply the CRD. In this case, we group the cows of the same location in a block. If each block has as many experimental units as the number of treatments and each and every treatment occurs once in a block, then the design is known as the randomised complete block design (RCBD). But in common parlance, it is called the randomised block design (RBD). Once the blocks are formed, treatments are allocated on the experimental units in each block randomly and independently.

Analysis of RBD is done with the help of two-way ANOVA with one observation per cell. It has been described in Unit 10. We briefly mention the main steps and formulae as follows:

**Step 1:** We first formulate the null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ). In RBD, we can test two hypotheses simultaneously: one for testing whether the effect of treatments or different levels of a treatment (factor) is significantly different and the other for testing whether the effect of different blocks is significantly different. If there are p treatments (or a treatment has p levels), we can take the null and alternative hypotheses for treatments as follows:

$H_{0A}$ : The average effect of all treatments is equal or the average effect of the treatments is not significantly different

$H_{1A}$ : The average effect of all treatments is not equal or the average effect of the treatments is significantly different

Symbolically,

$$H_{0A} : \alpha_1 = \alpha_2 = \dots = \alpha_p$$

$$H_{1A} : \text{At least one } \alpha_i \neq \alpha_j \ (i \neq j, 1, 2, \dots, p)$$

Similarly, if there are q blocks in an experiment, we can take the null and alternative hypotheses for blocks as follows:

$$H_{0B} : \beta_1 = \beta_2 = \dots = \beta_q$$

$H_{1B}$ : The average effect of all blocks is not equal or the average effect of blocks is significantly different

- Step 2:** We arrange the data of RBD in the format of two-way ANOVA, i.e., we arrange the data treatment-wise and block-wise.
- Step 3:** We analyse the RBD with the help of two-way ANOVA as explained in the screened box of Sec.11.4 of Lab Session 11 by taking blocks instead of factor B.
- Step 4:** If the null hypothesis of treatments is rejected, we do pair-wise comparison as explained in Sec. 11.3 of Lab Session 11.

### Steps in Excel

In Problem 2, we have to test whether the average effect of different chemicals on the strength of the cloth is significantly different and whether the effect of all bolts on the strength of cloth is the same. So we can formulate the null and alternative hypotheses for the chemicals and bolts (blocks) as follows:

$H_{0C}$ : The average effect of different chemicals on the strength of the cloth is not significantly different

$H_{1C}$ : The average effect of different chemicals on the strength of the cloth is significantly different (claim)

$H_{0B}$ : The average effect of all bolts on the strength of the cloth is the same (claim)

$H_{1B}$ : The average effect of all bolts on the strength of the cloth is not the same

The following steps are used to analysis RBD with the help of two-way ANOVA using the **Data Analysis ToolPak** in MS Excel 2007:

- Step 1:** We enter the data (given in Table 2) chemical-wise and bolt-wise in the format of two-way ANOVA in an Excel 2007 spreadsheet. We start by entering the heading of the data in Row 1, 2 and Column A in the Excel sheet and the data from Row 3 and Column B onwards as shown in Fig. 12.4.

	A	B	C	D	E	F	G
1	Chemical	Bolt					
2		1	2	3	4	5	6
3	C <sub>1</sub>	75	72	74	71	70	72
4	C <sub>2</sub>	73	72	70	70	68	70
5	C <sub>3</sub>	72	70	69	72	72	69
6	C <sub>4</sub>	70	68	70	68	66	67

Fig. 12.4: Partial screenshot of the spreadsheet for the given data.

- Step 2:** We repeat Steps 2 and 3 under the heading ‘Steps in Excel’ in Sec.11.4 of Lab Session 11 and obtain the results shown in Fig. 12.5.

A	B	C	D	E	F	G
<b>7 Anova: Two-Factor Without Replication</b>						
<b>8</b>						
9	<b>SUMMARY</b>	<b>Count</b>	<b>Sum</b>	<b>Average</b>	<b>Variance</b>	
10	C1	6	434	72.3333	3.4667	
11	C2	6	423	70.5000	3.1000	
12	C3	6	424	70.6667	2.2667	
13	C4	6	409	68.1667	2.5667	
14						
15	1	4	290	72.5	4.3333	
16	2	4	282	70.5	3.6667	
17	3	4	283	70.75	4.9167	
18	4	4	281	70.25	2.9167	
19	5	4	276	69	6.6667	
20	6	4	278	69.5	4.3333	
21						
22						
23	<b>ANOVA</b>					
24	<b>Source of Variation</b>	<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>P-value</b>
25	Rows	52.8333	3	17.6111	9.5482	0.0009
26	Columns	29.3333	5	5.8667	3.1807	0.0373
27	Error	27.6667	15	1.8444		4.5556
28						
29	<b>Total</b>	<b>109.833</b>	<b>23</b>			

Fig. 12.5

The explanation of the results is the same as given for two-way ANOVA in Step 3 under the heading ‘Steps in Excel’ in Sec.11.4 of Lab Session 11.

#### Decision using the critical region approach

Note that the calculated F value for different chemicals (rows) is 9.5482, which is greater than the corresponding critical value, 5.4170. It means that  $F_{\text{cal}}$  lies in the rejection region. So we reject the null hypothesis. Since the alternative hypothesis is the claim, we do not reject the claim. Hence, we conclude that there is not enough evidence against the claim. So we conclude that the average effect of different chemicals on the strength of the cloth is significantly different at 1% level of significance.

The calculated F value for bolts (columns) is 3.1807, which is less than the corresponding critical value, 4.5556. It means that  $F_{\text{cal}}$  lies in the non-rejection region. So we do not reject the null hypothesis. Since the null hypothesis is the claim, we do not reject the claim. Hence, we conclude that there is not enough evidence against the claim. So we may conclude that the effect of the bolts on the strength of the cloth is the same at 1% level of significance.

#### Decision using the p-value approach

Since the p-value for different chemicals ( $= 0.0009$ ) is less than  $\alpha$  ( $= 0.01$ ), we reject the null hypothesis.

Since the p-value for bolts of cloth ( $= 0.0373$ ) is greater than  $\alpha$  ( $= 0.01$ ), we do not reject the null hypothesis.

**Step 3:** Since the null hypothesis for chemicals is rejected, i.e., the average effect of different chemicals on the strength of cloth is significantly different, we do pair-wise comparison as explained in Step 4 under the heading ‘Steps in Excel’ in Sec. 11.3 of Lab Session 11. The final output is shown in Fig. 12.6.

	A	B	C	D	E	F	G	H	I
31	Comparison table								
32	Pair of treatments	$n_i$	$n_j$	$\bar{T}_i$	$\bar{T}_j$	$ \bar{T}_i - \bar{T}_j $	$t_{(15),0.005}$	CD	Inference
33	$C_1, C_2$	6	6	72.3333	70.5000	1.8333	2.9467	2.3105	Insignificant
34	$C_1, C_3$	6	6	72.3333	70.6667	1.6667	2.9467	2.3105	Insignificant
35	$C_1, C_4$	6	6	72.3333	68.1667	4.1667	2.9467	2.3105	Significant
36	$C_2, C_3$	6	6	70.5000	70.6667	0.1667	2.9467	2.3105	Insignificant
37	$C_2, C_4$	6	6	70.5000	68.1667	2.3333	2.9467	2.3105	Significant
38	$C_3, C_4$	6	6	70.6667	68.1667	2.5000	2.9467	2.3105	Significant

Fig. 12.6

From the comparison table shown in Fig 12.6, we observe that the effect of chemicals  $C_1$ ,  $C_2$  and  $C_3$  on the strength of the cloth is significantly different from that of chemical  $C_4$ . However, chemical  $C_1$  is not significantly different from chemicals  $C_2$  and  $C_3$ . Also, chemical  $C_2$  is not significantly different from chemical  $C_3$ .

## 12.5 LATIN SQUARE DESIGN

In Unit 11 of MST-005, you have learnt that the completely randomised design (CRD) is used when the experimental material is homogeneous and RBD is used when experimental material is not homogeneous and only one factor is likely to influence the effect of the treatments. The variation due to this factor is eliminated by grouping the experimental units in homogeneous groups called **blocks**. But in many situations, it is known that two factors may influence the effect of the treatments. For example, in the feeding experiment of cows, if the cows are uniform with respect to age and other factors but are of different breeds and from different locations, we cannot apply CRD and RBD. In such situations, Latin square design (LSD) is used to control the variability in two directions and we can group the cows of the same breed as a block (row) and the cows of the same location as another block (column). In Latin square design, we eliminate the variations due to these factors by grouping the experimental units in homogeneous groups (blocks) row-wise and column-wise. Once the blocks are formed, the treatments are allocated to the experimental units randomly and independently in such a way that each treatment occurs only once in each row and each column.

In LSD the number of rows, columns and treatments are equal.

The analysis of LSD has been described in Unit 11 of MST-005. We briefly mention the main steps and formulae as follows:

**Step 1:** We first formulate the null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ). In LSD, we can test three hypotheses simultaneously: one for testing whether the effect of different rows is significantly different, the second for testing whether the effect of different columns is significantly different, and the third for testing whether the effect of different treatments (or different levels of a treatment) is significantly different. If there are  $m$  rows in LSD we can take the null and alternative hypotheses for testing the effect of the rows as follows:

$H_{IR}$ : The average effect of all rows is equal, i.e., the average effect of all rows is not significantly different

or  $H_{0R}$  :  $\alpha_1 = \alpha_2 = \dots = \alpha_m$

$H_{IR}$ : The average effect of all rows is not equal, i.e., the average effect of rows is significantly different

Similarly, if there are  $m$  columns in LSD, we can take the null and alternative hypotheses for testing the effect of the columns as follows:

$H_{0C}$  : The average effect of all columns is equal, i.e., the average effect of all columns is not significantly different

or  $H_{0C} : \beta_1 = \beta_2 = \dots = \beta_m$

$H_{1C}$  : The average effect of all columns is not equal, i.e., the average effect of all columns is significantly different

Similarly, if there are  $m$  treatments (or  $m$  levels of a treatment) in LSD, we can take the null and alternative hypotheses for testing the effect of the treatments as follows:

$H_{1T}$  : The average effect of all treatments is equal, i.e., the average effect of all treatments is not significantly different

or  $H_{0T} : \tau_1 = \tau_2 = \dots = \tau_m$

$H_{1T}$  : The average effect of all treatments is not equal, i.e., the average effect of all treatments is significantly different

**Step 2:** We arrange the data of LSD in the format of two-way ANOVA, i.e., we arrange the data row-wise and column-wise.

**Step 3:** We calculate the correction factor (CF) and the raw sum of squares (RSS) using the formulae given below:

$$CF = \frac{G^2}{N} \quad \dots (1)$$

$$RSS = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m y_{ijk}^2 \quad \dots (2)$$

where  $G$  – the grand total, i.e., the sum of observations,

$N$  – the total number of observations, i.e.,  $N = m^2$ , and

$y_{ijk}$  – the observation in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column for the  $k^{\text{th}}$  treatment.

**Step 4:** We calculate the various sums of squares using the formulae given below:

$$TSS = RSS - CF \quad \dots (3)$$

$$SSR = \frac{1}{m} \sum_{j=1}^m y_{i..}^2 - CF \quad \dots (4)$$

$$SSC = \frac{1}{m} \sum_{j=1}^m y_{.j}^2 - CF \quad \dots (5)$$

$$SST = \frac{1}{m} \sum_{k=1}^m y_{..k}^2 - CF \quad \dots (6)$$

$$SSE = TSS - SSR - SSC - SST \quad \dots (7)$$

where  $y_{i..} = \sum_{j=1}^m \sum_{k=1}^m y_{ijk}$  – the sum of observations of the  $i^{\text{th}}$  row

$y_{.j} = \sum_{i=1}^m \sum_{k=1}^m y_{ijk}$  – the sum of observations of the  $j^{\text{th}}$  column

$y_{..k} = \sum_{i=1}^m \sum_{j=1}^m y_{ijk}$  – the sum of observations of the  $k^{\text{th}}$  treatment

**Step 5:** We calculate the degrees of freedom (df) for rows, columns, treatments and error as follows:

The degrees of freedom for rows =  $m - 1$

The degrees of freedom for columns =  $m - 1$

The degrees of freedom for treatments =  $m - 1$

The degrees of freedom for error =  $(m - 1)(m - 2)$

**Step 6:** We obtain the various mean sums of squares as follows:

$$\text{Mean sum of squares due to rows (MSSR)} = \frac{\text{SSR}}{m-1} \quad \dots (8)$$

$$\text{Mean sum of squares due to columns (MSSC)} = \frac{\text{SSC}}{m-1} \quad \dots (9)$$

$$\text{Mean sum of squares due to treatment (MSST)} = \frac{\text{SST}}{m-1} \quad \dots (10)$$

$$\text{Mean sum of squares due to error (MSSE)} = \frac{\text{SSE}}{(m-1)(m-2)} \quad \dots (11)$$

**Step 7:** We calculate the value of the test statistics using the formulae given below:

$$F_R = \frac{\text{MSSR}}{\text{MSSE}} \quad \dots (12)$$

$$F_C = \frac{\text{MSSC}}{\text{MSSE}} \quad \dots (13)$$

$$F_T = \frac{\text{MSST}}{\text{MSSE}} \quad \dots (14)$$

**Step 8:** We take the decision about the null hypothesis for rows, columns and treatments as explained in Step 7 of the screened box of Sec. 11.3 of Lab Session 11.

**Step 9:** If the null hypothesis of treatments is rejected, we do pair-wise comparison as explained in Sec. 11.3 of Lab Session 11.

## Steps in Excel

In Problem 3, we have to test whether the effect of soil fertility levels, water levels and different fertilisers are significantly different on the yield of the crop at 5% level of significance. So we can take the null and alternative hypotheses for the soil fertility levels, water levels and different fertilisers as follows:

$H_{0S}$  : The average effect of soil fertility levels on the yield of the crop is not significantly different

$H_{1S}$  : The average effect of soil fertility levels on the yield of the crop is significantly different (claim)

$H_{0W}$  : The average effect of water levels on the yield of the crop is not significantly different

$H_{1W}$  : The average effect of water levels on the yield of the crop is significantly different (claim)

$H_{0F}$  : The average effect of different fertilisers on the yield of the crop is not significantly different

$H_{1F}$  : The average effect of different fertilisers on the yield of the crop is significantly different (claim)

Since the **Data Analysis ToolPak** in MS Excel 2007 does not give direct analysis of LSD, we analyse the LSD using the following steps in MS Excel 2007:

**Step 1:** We enter the data given in Table 3 for soil fertility and water level in an Excel 2007 spreadsheet (Fig. 12.7). To show different fertilisers, we can use different colours. Here we use orange, blue, green and yellow colours for fertilisers A, B, C and D, respectively, as shown in Fig. 12.7.

	A Soil fertility level	B	C	D	E
1	Water level				
2		I	II	III	IV
3	1	14.55	9.45	14.70	2.85
4	2	5.20	12.10	17.50	9.55
5	3	2.70	15.40	12.00	16.50
6	4	12.45	20.85	4.75	14.45

Fig. 12.7: Partial screenshot of the spreadsheet for the given data.

**Step 2:** We calculate the sum of the observations for each soil fertility level (row), water level (column) and fertiliser as follows:

1. We first calculate the sum of the first soil fertility level 1 in Cell F2 by typing “=Sum(B3:E3)” and pressing **Entre** as shown in Fig. 12.8a and then dragging Cell F2 down up to Cell F6. We get the sum of soil fertility levels 2, 3 and 4 as shown in Fig 12.8b.

	A	B	C	D	E	F
2		I	II	III	IV	Total
3	1	14.55	9.45	14.70	2.85	41.55
4	2	5.20	12.10	17.50	9.55	44.35
5	3	2.70	15.40	12.00	16.50	46.60
6	4	12.45	20.85	4.75	14.45	52.50

Fig. 12.8

2. Similarly, we calculate the sum of the observations for different water levels (I, II, III and IV) in Cells B7, C7, D7 and E7 as shown in Figs. 12.9a and b.

	A	B	C	D	E	F
1		I	II	III	IV	
2		14.55	9.45	14.70	2.85	41.55
3	2	5.20	12.10	17.50	9.55	44.35
4	3	2.70	15.40	12.00	16.50	46.60
5	4	12.45	20.85	4.75	14.45	52.50
6	Total	34.90				
7	Total	57.80	48.95	43.35		
8						

Fig. 12.9

3. We now calculate the sum of the observations of different fertilisers. Here we show the cells receiving different fertilisers by different colours. So we calculate the sum of observations of fertiliser A by selecting orange cells, i.e., by typing “=E3+C4+B5+D6” in Cell B8 and pressing **Enter**. Then we get the output shown in Fig. 12.10a. Similarly, we calculate the sum of observations of fertilisers B, C and D in Cells C8, D8 and E8, respectively as shown in Fig. 12.10b.

The figure consists of two separate Excel spreadsheets, labeled (a) and (b). Both spreadsheets have columns A through F and rows 1 through 9. Column A contains labels 1, 2, 3, 4, Total, and Total. Column B contains values 14.55, 5.20, 2.70, 12.45, 34.90, and 22.40. Column C contains values 9.45, 12.10, 15.40, 20.85, 57.80, and 88.90. Column D contains values 14.70, 17.50, 12.00, 4.75, 48.95, and 74.05. Column E contains values 2.85, 9.55, 16.50, 14.45, 43.35, and 60.85. Row 1 is a header row with column labels A through F. Row 9 is a footer row labeled 'Total' across all columns except A. The formula bar at the top of each spreadsheet shows '=E3+C4+B5+D6'. The cells containing the formulas (B8, C8, D8, E8) are highlighted in orange.

	A	B	C	D	E	F
2		I	II	III	IV	Total
3	1	14.55	9.45	14.70	2.85	41.55
4	2	5.20	12.10	17.50	9.55	44.35
5	3	2.70	15.40	12.00	16.50	46.60
6	4	12.45	20.85	4.75	14.45	52.50
7	Total	34.90	57.80	48.95	43.35	
8	Total	22.40				
9						

	A	B	C	D	E	F
2		I	II	III	IV	Total
3	1	14.55	9.45	14.70	2.85	41.55
4	2	5.20	12.10	17.50	9.55	44.35
5	3	2.70	15.40	12.00	16.50	46.60
6	4	12.45	20.85	4.75	14.45	52.50
7	Total	34.90	57.80	48.95	43.35	
8	Total	22.40	88.90	74.05	60.85	
9						

Fig. 12.10

**Step 3:** We calculate the sum (G) of all observations given in Cells B3:E6 in Cell B9 by typing “=Sum(B3:E6)” and pressing **Enter** as shown in Fig. 12.11.

The figure shows a single Excel spreadsheet with columns A through E and rows 8 and 9. Column A contains labels Total and G. Column B contains values 22.40 and 185. Column C contains values 43.45, 57.25, and 61.90. Column D is empty. Column E is empty. Row 8 is a header row with column labels A through E. Row 9 is a footer row labeled 'G' across all columns except A. The formula bar at the top shows '=SUM(B3:E6)'. The cell containing the formula (B9) is highlighted in orange.

	A	B	C	D	E
8	Total (y <sub>..k</sub> )	22.40	43.45	57.25	61.90
9	G	185			

Fig. 12.11

**Step 4:** We calculate the correction factor ( $CF = G^2/N$ ) in Cell B10 by typing “=B9\*B9/16” and pressing **Enter**. The result is shown in Fig. 12.12.

The figure shows a single Excel spreadsheet with columns A through E and rows 9 and 10. Column A contains labels G and CF. Column B contains values 185 and 2139.0625. Column C is empty. Column D is empty. Column E is empty. Row 9 is a header row with column labels A through E. Row 10 is a footer row labeled 'CF' across all columns except A. The formula bar at the top shows '=B9\*B9/16'. The cell containing the formula (B10) is highlighted in orange.

	A	B	C	D	E
9	G	185			
10	CF	2139.0625			

Fig. 12.12

**Step 5:** We now calculate the raw sum of squares ( $RSS = \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m y_{ijk}^2$ ), i.e.,

the sum of squares of all observations given in Cells B3:E6 as follows:

- We select Cell B11,

2. Then we click on **Formulas** → **Math & Trig**→ **Sumsq** function as shown in Fig. 12.13.

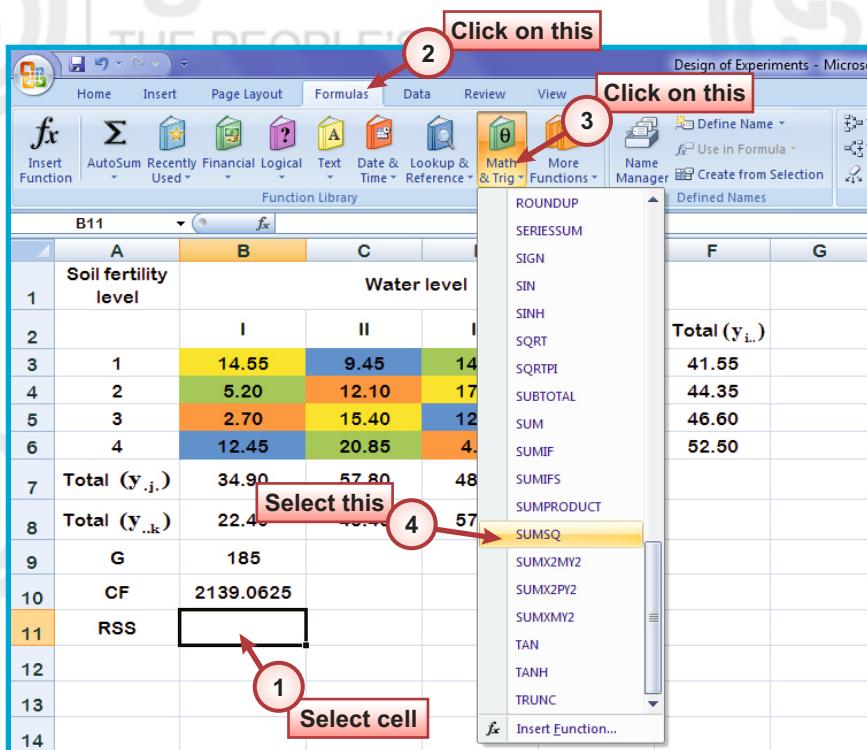


Fig. 12.13

3. A new dialog box is opened. This dialog box requires the **Number 1** for which we calculate the sum of squares. So we select Cells B3:E6 in this box and then click on **OK** (see Fig. 12.14a). This function gives the value of RSS as shown in Fig. 12.14b.

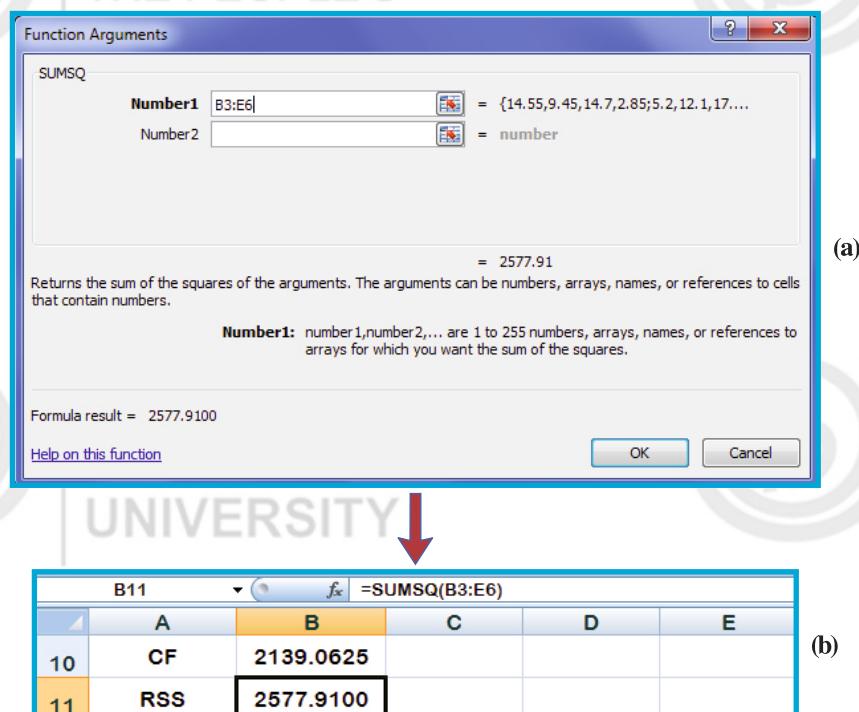


Fig. 12.14

**Step 6:** We calculate  $\sum_{i=1}^m y_{i..}^2$ ,  $\sum_{j=1}^m y_{.j}^2$  and  $\sum_{k=1}^m y_{..k}^2$  in Cells B12, B13 and B14,

respectively, as explained in Step 5. The outputs are shown in Figs. 12.15a, 12.15b and 12.15c, respectively.

(a)

	A	B	C	D	E
11	RSS	2577.9100			
12	$\sum y_{i..}^2$	8621.1350			

(b)

	A	B	C	D	E
12	$\sum y_{i..}^2$	8621.1350			
13	$\sum y_{..j}^2$	8834.1750			

(c)

	A	B	C	D	E
13	$\sum y_{..j}^2$	8834.1750			
14	$\sum y_{..k}^2$	9498.8350			

Fig. 12.15

**Step 7:** We now prepare the ANOVA table as shown in Fig. 12.16.

	A	B	C	D	E	F
16	ANOVA					
17	Source of variation	SS	df	MSS	$F_{cal}$	$F_{crit}$
18	Soil fertility					
19	Water level					
20	Fertiliser					
21	Error					
22	Total					

Fig. 12.16

**Step 8:** We calculate the sum of squares (SS) for total (TSS), soil fertility (SSS), water level (SSW), fertiliser (SSF) and error in Cells B22, B18, B19, B20 and B21 by typing “=B11-B10”, “=B12/4-B10”, “=B13/4-B10”, “=B14/4-B10” and “=B22-B18-B19-B20”, respectively, and pressing **Enter**. The results are shown in Fig. 12.17.

	A	B	C	D	E	F
16	ANOVA					
17	Source of variation	SS	df	MSS	$F_{cal}$	$F_{crit}$
18	Soil fertility	16.2213				
19	Water level	69.4813				
20	Fertiliser	235.6463				
21	Error	117.4987				
22	Total	438.8475				

Fig. 12.17

**Step 9:** We type the degrees of freedom for soil fertility, water level, fertiliser and error in Cells C18, C19, C20 and C21, respectively. The results are shown in Fig. 12.18.

	A	B	C	D	E	F
16	ANOVA					
17	Source of variation	SS	df	MSS	$F_{cal}$	$F_{crit}$
18	Soil fertility	16.2213	3			
19	Water level	69.4813	3			
20	Fertiliser	235.6463	3			
21	Error	117.4987	6			
22	Total	438.8475				

Fig. 12.18

**Step 10:** We calculate the mean sum of squares (MSS) for soil fertility, water level, fertiliser and error in Cells D18, D19, D20 and D21 by typing “=B18/C18”, “=B19/C19”, “=B20/C20” and “=B21/C21”, respectively, and pressing **Enter**. The results are shown in Fig. 12.19.

	A	B	C	D	E	F
16	ANOVA					
17	Source of variation	SS	df	MSS	$F_{cal}$	$F_{crit}$
18	Soil fertility	16.2213	3	5.4071		
19	Water level	69.4813	3	23.1604		
20	Fertiliser	235.6463	3	78.5488		
21	Error	117.4987	6	19.5831		
22	Total	438.8475				

Fig. 12.19

**Step 11:** We calculate the F value ( $F_{cal}$ ) for soil fertility, water level and fertiliser in Cells E18, E19 and E20 by typing “=D18/D21”, “=D19/D21” and “=D20/D21”, respectively and pressing **Enter**. The results are shown in Fig. 12.20.

	A	B	C	D	E	F
16	ANOVA					
17	Source of variation	SS	df	MSS	$F_{cal}$	$F_{crit}$
18	Soil fertility	16.2213	3	5.4071	0.2761	
19	Water level	69.4813	3	23.1604	1.1827	
20	Fertiliser	235.6463	3	78.5488	4.0110	
21	Error	117.4987	6	19.5831		
22	Total	438.8475				

Fig. 12.20

**Step 12:** We finally obtain the critical value of F for soil fertility, water level and fertiliser in Cells F18, F19 and F20 as explained in Step 3 under the heading ‘Steps in Excel’ in Sec. 10.7 of Lab Session 10. It means that we calculate the critical value for soil fertility as follows:

1. We select Cell F18,
2. Then we click on the **Formulas → More Functions → Statistical → Finv** function,
3. A new dialog box opens (Fig. 12.21a). We type the value of level of significance  $\alpha = 0.05$  in the box **probability**. We select the Cells C18 and C21 in the box **Deg\_freedom 1** and **Deg\_freedom 2**, respectively, as shown in Fig. 12.21a.
4. Then we click on **OK**. This function gives the output shown in Fig. 12.21b.

Similarly, we obtain the critical values for the water level and the fertiliser. The results are shown in Fig. 12.21c.

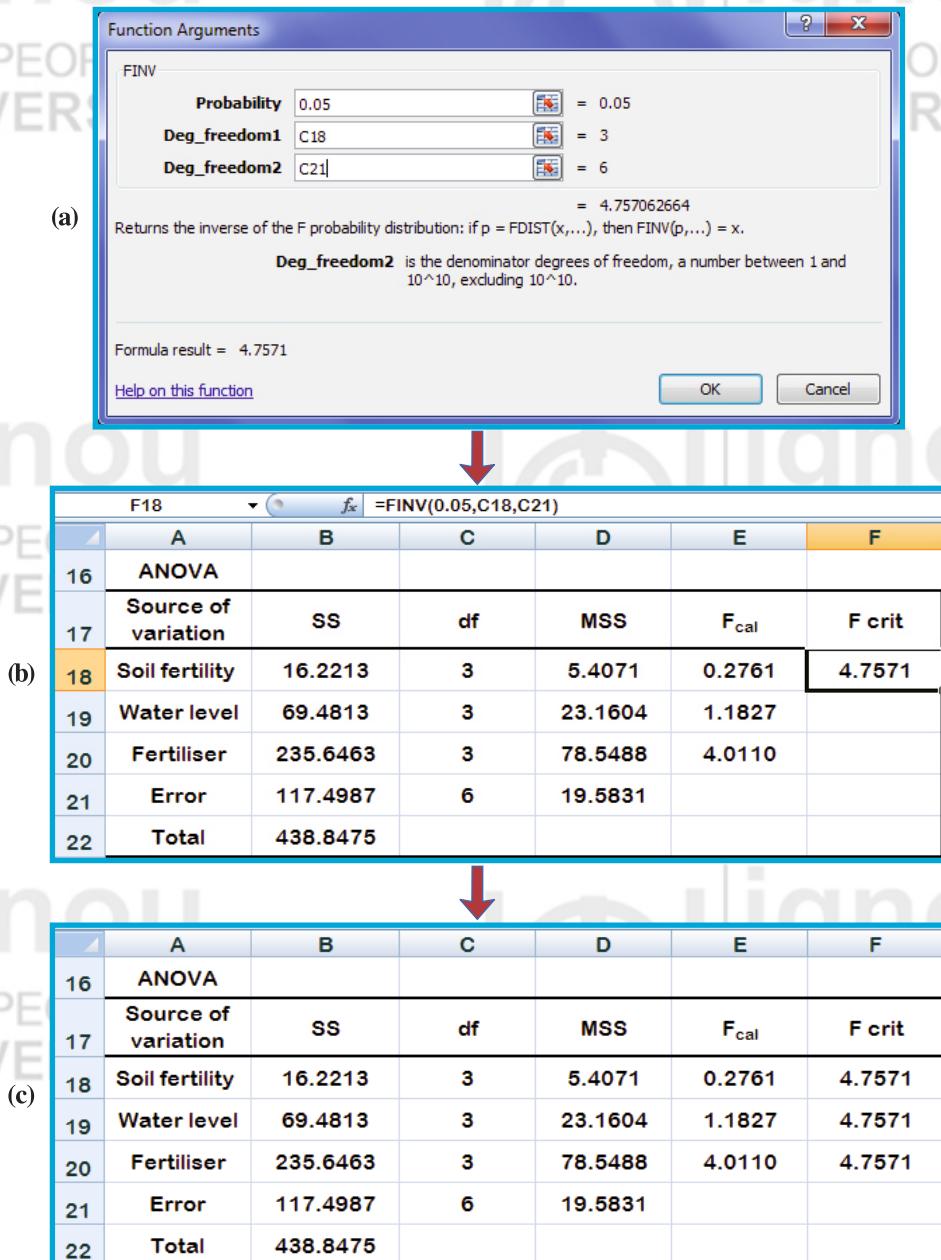


Fig. 12.21

### Decision using the critical region approach

Note that the calculated value for soil fertility is 0.2761, which is less than the corresponding critical value, 4.7571. It means that it lies in the non-rejection region. So we do not reject the null hypothesis. Since the alternative hypothesis is the claim, we reject the claim. Hence, we conclude that there is enough evidence against the claim. So we may conclude that the effect of the soil fertility levels on the yield of the crop is not significantly different at 5% level of significance.

The calculated value for the water levels is 1.1827, which is less than the corresponding critical value, 4.7571. It means that it lies in the non-rejection region. So we do not reject the null hypothesis. Since the alternative hypothesis is the claim, we reject the claim. Hence, we conclude that there is enough evidence against the claim. So we may conclude that the effect of the water levels on the yield of the crop is not significantly different at 5% level of significance.

The calculated value for the fertilisers is 4.0110, which is less than the corresponding critical value, 4.7571. It means that it lies in the non-rejection region. So we do not reject the null hypothesis. Since the alternative hypothesis is the claim, we reject the claim. Hence, we conclude that there is enough evidence against the claim. So we may conclude that the effect of different fertilisers on the yield of the crop is not significantly different at 5% level of significance.

**Note:** In this problem, since the null hypothesis for fertilisers is not rejected, we do not do pair-wise comparison.



## Activity

For the following exercises, analyse the data with the help of MS Excel 2007 and interpret the results:

- A1) Example 1 given in Unit 9 of MST-005.
- A2) Exercise E1 given in Unit 9 of MST-005.
- A3) Examples 1 and 2 given in Unit 10 of MST-005.
- A4) Exercise E1 given in Unit 10 of MST-005.
- A5) Example 1 given in Unit 10 of MST-005.
- A6) Exercise E1 given in Unit 10 of MST-005.

Match the results with the manual computation of data carried out in Units 9 and 10 of MST-005.



## Continuous Assessment 12

1. A leading shirt manufacturer has 500 showrooms across the country. The company would like to test whether the average difference of sales between salespersons is significant. For ascertaining the sales of different salespersons, the company has adopted the practice of retaining one salesperson for three months in a showroom. The company randomly selects four salespersons (A, B, C and D) and randomly assigns them to different showrooms. The data for average sales (in thousand rupees) of the four salespersons in different showrooms are given in Table 4.

Table 4: Sales generated by four salespersons

85(A)	92(C)	80(A)	74(B)
86(C)	90(B)	75(B)	86(D)
90(D)	100(D)	92(A)	80(C)
80(A)	95(B)	82(C)	75(C)
70(A)	95(B)	74(D)	80(D)
84(B)	90(D)	78(C)	95(A)

Assuming that the sales of each salesperson are normally distributed with approximately equal variances, test whether the difference of average sales between salespersons is significant at 1% level of significance.

2. An agricultural scientist wishes to test the effect of four types of seeds (A, B, C and D) on the yield of a crop. Since soil fertility could be a potential source of variability, the scientist decides to use a randomised block design. She divides the experimental field into homogeneous blocks of approximately the same soil fertility and divides each block into four experimental units. Then she randomly allocates each seed to each block randomly. The yields obtained are shown in the following table:

**Table 5: Yields of the crop**

<b>Block I</b>	42(A)	44(B)	40(C)	35(D)
<b>Block II</b>	46(B)	38(D)	44(A)	45(C)
<b>Block III</b>	40(A)	43(C)	48(B)	42(D)
<b>Block IV</b>	40(D)	46(A)	38(C)	42(B)
<b>Block V</b>	42(C)	47(A)	40(D)	46(B)

Analyse the design at 5% level of significance by assuming that the effect of each seed on the yield of the crop is normally distributed with approximately equal variances and the effect of each block on the yield of the crop is normally distributed with approximately equal variances. Test whether the effect of the four types of seeds on the crop yield is the same.

3. A company wishes to test the efficiency of five operators (A, B, C, D and E) in terms of the number of units produced per day. The analyst of the company knows that there might be variability in the machines and days. So she decides to use a Latin square design by considering machines and days as blocks. She obtains the following result:

**Table 6: The number of units produced**

<b>Day</b>	<b>Machine</b>				
	<b>I</b>	<b>II</b>	<b>III</b>	<b>IV</b>	<b>V</b>
1	11(A)	26(E)	20(B)	15(A)	12(C)
2	5(B)	22(D)	10(A)	18(C)	24(E)
3	21(D)	24(B)	15(C)	30(E)	10(A)
4	12(E)	15(C)	27(D)	16(A)	15(B)
5	15(C)	12(A)	23(E)	26(B)	29(D)

Test at 5% level of significance:

- i) whether the effect of the operators on the production per day is significantly different by assuming that the effect of each operator is normally distributed with approximately equal variances.
- ii) whether the effect of the machines on the production per day is significantly different by assuming that the effect of each machine is normally distributed with approximately equal variances.
- iii) whether the effect of the days on the production per day is significantly different by assuming that the effect of each day is normally distributed with approximately equal variances.



## Home Work: Do It Yourself

- 1) Follow the steps explained in Secs. 12.3, 12.4 and 12.5 to apply the tests on the data of Tables 1, 2 and 3. Take the final screenshots and keep them in your record book.
- 2) Develop the spreadsheets for the exercises of “Continuous Assessment 12” as explained in this lab session. Take screenshots of the final spreadsheets.
- 3) **Do not forget** to keep all screenshots in your record book as these will contribute to your continuous assessment in the Laboratory.





Dear Learner,

While performing the Lab Sessions of MSTL -001, you may have found certain portions of the sessions difficult to perform. We wish to know your difficulties and suggestions in order to improve the course. Therefore, we request you to fill and send us the following feedback form, which pertains to this course. Just fill in the space provided or tick the most relevant option. If you find that the space provided is insufficient, kindly use a separate sheet.

*Please mail to:*

**The Course Coordinator (MSTL-001)**

**Room No. 106**

**School of Sciences,**

**Block D, New Academic Complex,**

**IGNOU, Maidan Garhi**

**New Delhi – 110068**

### **FEEDBACK FORM**

1. Name: ..... 2. Male/Female: ..... 3. Age: .....

4. Enrolment No: ..... 5. Email Address: .....

6. Mobile No: ..... 6. Employed/Unemployed: .....

7. Regional Centre Code and Name: .....

8. Study Centre Code and Name: .....

9. How many hours did you spend for performing all lab sessions?

Part No.	A					B			C		D	
Session No.	1	2	3	4	5	6	7	8	9	10	11	12
No. of Hours												

10. How do you find the relevance of lab sessions for theory courses?

- i) Excellent    ii) Good    iii) Adequate    iv) Poor    v) Very poor

11. Did you enjoy practicing with MS Excel 2007?                          Yes/No

12. Do you have a personal computer?                          Yes/No

13. Where do you access the internet?

- i) At home    ii) At internet cafe    iii) At work place    iv) Other .....

- 14) Put a tick mark on what you think is correct, based on your experience with the course from amongst the options given below:

Items	Excellent	Good	Adequate	Poor	Very Poor	Give Specific Examples, if Poor or Very Poor
Quality of Presentation						
Language and Style						
Screenshots Used						
Conceptual Clarity						
Activities						
Continuous Assessment Exercises						

- 15) **Specific Feedback:** If you have found some sessions of the course to be difficult and not clear, please mention the page numbers and what is not clear about each lab session. Use additional pages, if need be.

- 16) Your overall comments and suggestions on the course (Please write the comments on the course after studying it).

# UNIVERSITY