

# 2

## Some Basic Experimental Designs

---

### 2.1 Introduction

An experimental design is essentially a rule that determines the assignment of treatments to the experimental units keeping in mind the principles of randomization, replication and local control. The experiments, however, differ from each other greatly in many respects, depending upon the variability in the experimental units and how it is taken care of. Generally the process of randomization of the experimental units depends upon the way the variability present in the experimental units is accounted for. This then dictates what type of design is used in a given experimental situation. In some experimental situations, a naïve design, generated keeping in mind the experimental conditions and practical considerations, helps answering the objectives of the experiment. Nonetheless, there are some basic (or standard) designs that are used frequently by the experimenters because of the ease in running these experiments. The purpose of this Chapter is to describe such designs.

We begin with a Completely Randomized Design (CRD), which uses the principles of randomization and replication. This design is used when there are strong reasons to believe that there is no variability in the experimental units. This will be followed by Randomized Complete Block (RCB) Design and Latin Square Design, in which all the three principles of randomization, replication and local control are applied. RCB design and Latin Square Design are used when there are one and two sources of variability, respectively present in the experimental units. In both these designs, the treatment replications are equal. There is flexibility in the choice of number of replications in RCB design, but in case of Latin square design, the replication number is equal to the number of treatments. As would be seen later in this Chapter and in other Chapters as well, it is simply the change in the randomization procedure of the treatments to the experimental units that gives rise to different designs. There could, however, be more sources of variability present in the experimental units and the randomization would be done accordingly. It will become obvious through different Chapters that the randomization and control of variability through grouping(s) of the experimental units are related to each other and help in controlling the experimental error.

### 2.2 Completely randomized design

Consider an experimental situation where the experimenter is interested (a) in comparing four grazing systems (treatments), viz., *rotational*, *deferred rotational*, *continuous* and *cut and carry*, and (b) to study the effect of the grazing systems on the body weight of the animals. Suppose that 16 animals are available for conducting the experiment. Suppose further that the choice of 16 animals is such that they do not contribute to the variability in the final body

weights of the animals after being subjected to the grazing systems. In other words, it is assumed that the experimental units (subjects or animals) do not contribute to the variability in the data and the only explainable part of variability present in the data is because of the four different grazing systems. The unexplained part of variability is the experimental error. An easy way of running this experiment is to allocate the 4 grazing systems randomly to the 16 animals such that each grazing system is received by four animals. However, it is not necessary to have equal replication of grazing systems. One can have unequal replication of the treatments as well. But as far as possible, it is better to have equal, or as equal as possible, replications of the treatments. If that be so, then the replications would differ by at most one. This is known as a design for zero-way elimination of heterogeneity.

A zero-way heterogeneity setting design or an unblocked design or a CRD is the simplest design in which only two principles of design of experiments *viz.* randomization and replication are used. There is no use of local control here, since the experimental units are assumed to be homogeneous. The only identifiable cause of variability is the treatments and the remaining part of the variability is the experimental error.

To make the exposition general, suppose that there are  $\nu$  treatments and  $n$  homogeneous experimental units. The  $\nu$  treatments are allotted at random to the  $n$  experimental units. Let the  $i$ th treatment be replicated  $r_i$  times ( $i = 1, 2, \dots, \nu$ ) such that  $\sum_{i=1}^{\nu} r_i = n$ . Normally the number of replications for different treatments should be equal as it ensures equal precision of estimates of linear functions of treatment effects. The average replication number is then  $n/\nu$ , which will be a positive integer if  $\nu$  divides  $n$ . The actual number of replications of treatments is, however, determined by the availability of experimental resources and the requirement of precision and sensitivity of comparisons. If the experimental material for some treatments is available in limited quantities, the number of replications of these treatments is reduced. If the estimates of certain treatment effects are required with more precision, the number of replications of such treatments is increased.

### 2.2.1 Randomization

There are several methods of random allocation of treatments to the experimental units. The  $\nu$  treatments are first assigned numbers (or labels) randomly from 1 to  $\nu$ . The  $n$  experimental units are also numbered randomly. One method of randomization uses the random number tables. Any column (or columns) of a randomly opened page of a random number table is taken. If  $\nu$  is a one-digit number, then only one column is consulted digit by digit. If  $\nu$  is a two-digit number, then two columns (or two-digit random numbers) are consulted. All numbers greater than  $\nu$  and zero, are ignored.

Let the first number chosen be  $n_1$ ; then the treatment numbered  $n_1$  is allotted to the first unit. If the second number is  $n_2$  which may or may not be equal to  $n_1$ , then the treatment numbered  $n_2$  is allotted to the second unit. This procedure is continued. When the  $i$ th treatment number has occurred  $r_i$  times, ( $i = 1, 2, \dots, \nu$ ) this treatment is ignored subsequently. This process terminates when all the units are exhausted.

One drawback of the above procedure is that sometimes a very large number of random numbers may have to be ignored because they are greater than  $\nu$ . It may even happen that the random number table is exhausted before the allocation is complete. To avoid this difficulty the following procedure is adopted.

Let  $\nu$  be an  $s$ -digit number. Choose  $P$  as the highest  $s$ -digit number divisible by  $\nu$ . For instance, if  $\nu = 13$ , then  $P = 91$ ; when  $\nu = 31$ , then  $P = 93$ ; when  $\nu = 123$ , then  $P = 984$ . All numbers greater than  $P$  and zero are ignored. If a selected random number is less than  $\nu$ , then it is used as such. If it is greater than or equal to  $\nu$ , then it is divided by  $\nu$  and the remainder is taken to be the random number and used for allotting treatment to experimental unit. When a number is divisible by  $\nu$  (i.e., the remainder is zero), then the random number is  $\nu$ . For example, assume that  $\nu = 123$  and the random number drawn is 991. This number would be rejected because this is greater than 984. If the random number drawn is 95, then the treatment labeled 95 is allotted to that experimental unit. Further, if the random number selected is 567, then dividing 567 by 123 would leave the remainder as 75. So treatment labeled 75 is allotted to that experimental unit. Further, if the random number selected is 615, then dividing 615 by 123 would leave the remainder as zero. In this case, treatment labeled as 123 is allotted to that experimental unit.

### Alternative methods of random allocation

If random number tables are not available, treatments can be allotted by drawing *lots* as explained in the sequel. However, these procedures may not help generate strictly random numbers. So these procedures need to be adopted with caution.

The number of the  $i$ th treatment is written on  $r_i$  pieces of papers ( $i = 1, 2, \dots, \nu$ ). The  $\sum_{i=1}^{\nu} r_i = n$  pieces of papers are then folded individually so that the numbers written on them are not visible. These papers are then drawn one by one at random. Before each draw the slips are thoroughly shuffled. The treatment that is drawn at the  $t$ th draw is allotted to the  $t$ th unit ( $t = 1, 2, \dots, n$ )

Random allocation is also possible by using a fair coin. Let there be 5 treatments and 20 experimental units. Each treatment is to be replicated four times. Suppose that the experimental units are labeled by numbers from 1 to 20 randomly.

When a coin is tossed, there are two possible outcomes; either head or tail appears. Denote the “head” by H and the “tail” by T. When the coin is tossed twice, there are four possible outcomes; these are HH, HT, TH or TT. Similarly, when the coin is flipped three times, there are eight possible outcomes; HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. This can be easily generalized to  $n$  flippings of the coin.

The 5 treatments are now identified not by serial numbers as earlier but by any five of the above eight possible outcomes obtainable by flipping a coin three times. If any of the remaining three outcomes, say THT, TTH and TTT appear, no treatment is selected for allotment and the coin is again flipped thrice.

A coin is now thrown three times and the outcome is noted. If the outcome is any of the five outcomes HHH, HHT, HTH, HTT, THH, the treatment labeled by it is allotted to the first experimental unit. If the event happened is any of the three, THT, TTH, TTT, it is ignored. The coin is again tossed three times and this event is used to select a treatment for the second experimental unit. If the same outcome appears more than once, do not reject it until the number of times it has appeared equals the number of replications of the treatment it represents. This process is continued till all the experimental units are exhausted.

It may be worthwhile mentioning here that the labels are also allotted randomly to all the treatments. This would hold everywhere, whether mentioned or not.

The linear model in this case is

$$\text{Expected response} = \text{general mean} + \text{effect of treatments.}$$

Since there is no source of variation in the experimental units, the model does not contain the effect due to experimental units.

This can also be written as

$$\text{response} = \text{general mean} + \text{treatments effect} + \text{error,}$$

where the errors are independently distributed as normal variate with zero mean and constant variance  $\sigma^2$ . The partitioning of the total variability in this case is

<b>Source of variation</b>
Due to model
Error
Total

The component “due to model” can be partitioned as

<b>Source of variation</b>
Due to model
Due to treatments

### 2.2.2 Analysis of CRD

This design provides a one-way classified data according to levels of a single factor, the single factor being the treatments with  $v$  levels. Since no variability is expected from the experimental units, the only identifiable source of variability is the treatments. We then have the following linear model:

$$y_{ij} = \mu + \tau_i + e_{ij}, \quad i = 1, \dots, v; j = 1, \dots, r_i$$

$$E(e_{ij}) = 0, \quad \text{Cov}(e_{ij}, e_{kl}) = \begin{cases} \sigma^2, & \text{if } i = k, j = l \\ 0, & \text{otherwise} \end{cases}$$

where the random variable  $y_{ij}$  is the observation recorded on the  $j$ th replicate of the  $i$ th treatment,  $\mu$  is the general mean,  $\tau_i$  is the fixed effect of the  $i$ th treatment and  $e_{ij}$  is the random error component associated with the  $(i, j)$ th observation,  $i = 1, 2, \dots, v$ ;  $j = 1, 2, \dots, r_i$ . These are assumed to be distributed independently and normally with zero mean and constant variance  $\sigma^2$ . We also assume that the replication of the  $i$ th treatment is  $r_i$ ,  $i = 1, 2, \dots, v$  and  $\sum_{i=1}^v r_i = n$ .

Let us define the following:

$$\text{Treatment totals, } T_i = \sum_{j=1}^{r_i} y_{ij} \quad \forall i = 1, 2, \dots, v, \text{ and Grand total as } G = \sum_{i=1}^v T_i = \sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}.$$

The following formulae can be employed for analysis of variance:

$$\text{Correction factor (CF)} = \frac{G^2}{n}$$

$$\text{Total sum of squares (SS)} = \sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}^2 - CF$$

$$\text{Sum of squares due to treatments (SST)} = \sum_{i=1}^v \frac{T_i^2}{r_i} - CF$$

$$\text{Error sum of squares (SSE)} = \sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}^2 - \sum_{i=1}^v \frac{T_i^2}{r_i} = \left( \sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}^2 - CF \right) - \left( \sum_{i=1}^v \frac{T_i^2}{r_i} - CF \right)$$

$$= \text{Total SS} - \text{Treatment SS.}$$

The interest of the experimenter is in testing the null hypothesis:  $H_0 : \tau_1 = \tau_2 = \dots = \tau_v = \tau$  (say) against the alternative that  $\tau_l \neq \tau_i$ ,  $l \neq i = 1, 2, \dots, v$  for at least one pair of treatment effects, say  $\tau_l$  and  $\tau_i$ . For testing this hypothesis, we set up the analysis of variance Table 2.1.

**Table 2.1: ANOVA table in CRD**

Source	DF	SS	MS = SS/DF	F
Treatments	$v - 1$	$SST = \sum_{i=1}^v \frac{T_i^2}{r_i} - CF$	$\frac{SST}{(v-1)} = s_t^2$	$\frac{s_t^2}{s_e^2}$
Error	$n - v$	$SSE = \sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}^2 - \sum_{i=1}^v \frac{T_i^2}{r_i}$	$\frac{SSE}{(n-v)} = s_e^2$	
Corrected Total	$n - 1$	$\sum_{i=1}^v \sum_{j=1}^{r_i} y_{ij}^2 - CF$		

If the calculated value of  $F$  is greater than the table value of  $F_{\alpha; (v-1), (n-v)}$  at  $\alpha$  level of significance and  $(v - 1)$ ,  $(n - v)$  degrees of freedom, then the null hypothesis  $H_0$  is rejected at  $\alpha$  level of significance and it can be concluded that equality of all the treatment effects does not hold. In that case, the researcher has no knowledge about the treatment effects except that there is at least one pair of treatments that differs significantly from each other. In that case the researcher has to go for the computation of least significant difference (LSD) or other multiple comparison procedures as explained in Annexure-III to make pairwise treatment comparisons.

It may be seen here that the unbiased estimator of  $\sigma^2$  is  $\hat{\sigma}^2 = s_e^2$ .

Further, all the elementary treatment contrasts (or treatment contrasts for pairwise treatment comparisons) are estimable through the design. The best linear unbiased estimator (BLUE) of any treatment contrast  $d_{il} = \tau_i - \tau_l$  is

$$\hat{d}_{il} = \frac{T_i}{r_i} - \frac{T_l}{r_l}, \forall i \neq l = 1, 2, \dots, v.$$

The variance of  $\hat{d}_{il}$  is  $\text{Var}(\hat{d}_{il}) = \sigma^2 \left( \frac{1}{r_i} + \frac{1}{r_l} \right)$ . The estimated standard error of the estimated difference

between the  $i$ th and  $l$ th treatment effects is  $\hat{\text{SE}}(\hat{d}_{il}) = s_e \left( \frac{1}{r_i} + \frac{1}{r_l} \right)^{1/2}$ .

The least significant difference (LSD) is given as  $\text{LSD} = \hat{\text{SE}}(\hat{d}_{il}) \times t_{\alpha/2; \text{errorDF}}$ .

Here  $t_{\alpha; \text{error DF}}$  denotes the value of Student's  $t$  at  $\alpha$  level of significance and error degrees of freedom. The treatment means are given by  $\bar{T}_i = \frac{T_i}{r_i} \forall i = 1, 2, \dots, v$ . The pairwise comparison of treatment effects can be made by comparing the difference between any two treatment means with the LSD. Any two treatment effects are said to differ significantly if the difference of their means is larger than the LSD.

### 2.2.3 Example 1

An experiment was conducted in Rabi season on a variety of tomato during 2010-11 with 5 treatments of integrated nutrient management *viz.* Trt1 ~ farmers' practice (2.5 tonnes farmyard manure/ha), Trt2 ~ recommended dose of fertilizers (NPK 120:75:100), Trt3 ~ 50% recommended dose of fertilizers + vermin-compost 5 tonnes/ha, Trt4 ~ 50% recommended dose of fertilizers + vermin-compost 10 tonnes/ha and Trt5 ~ 50% recommended dose of fertilizers + vermin-compost 2.5 tonnes/ha + farmyard manures 5 tonnes/ha. The objective of the experiment was to find out the most appropriate integrated nutrient management system for tomato. The experiment was conducted using a completely randomized design and the dry matter accumulation (gm/plant) was recorded after the experiment was over. Table 2.2 gives the replicated data on dry matter accumulation in g/plant for each treatment:

**Table 2.2: Dry matter accumulation in g/plant**

Tr1	Tr2	Tr3	Tr4	Tr5
108.2	225.2	176.5	201.3	214.3
112.7	226.4	195.2	183.6	226.2
116.8	135.2	188.4	197.5	215.0
106.8	227.5	190.3	186.1	230.6
117.9	218.2	210.3	188.6	212.6
	229.1	195.1	210.4	230.4
				227.6
				228.3

In the sequel the data are analyzed to identify the best integrated nutrient management system.

### 2.2.4 Procedure and Calculations

The inference problem being solved here is the testing of the following null hypothesis:  $H_0: \tau_1 = \tau_2 = \dots = \tau_i = \dots = \tau_v = \tau$  (say) against the alternative hypothesis  $H_1$ : at least two of the  $\tau_i$ 's are different. In the example,  $v = 5$ .

First compute the following totals:

Treatment totals

$$T_1 = 108.2 + 112.7 + 116.8 + 106.8 + 117.9 = 562.4$$

$$T_2 = 225.2 + 226.4 + 135.2 + 227.5 + 218.2 + 229.1 = 1261.6$$

$$T_3 = 176.5 + 195.2 + 188.4 + 190.3 + 210.3 + 195.1 = 1155.8$$

$$T_4 = 201.3 + 183.6 + 197.5 + 186.1 + 188.6 + 210.4 = 1167.5$$

$$T_5 = 214.3 + 226.2 + 215 + 230.6 + 212.6 + 230.4 + 227.6 + 228.3 = 1785.0$$

$$\text{Gross total (G)} = T_1 + T_2 + T_3 + T_4 + T_5 = 562.4 + 1261.6 + 1155.8 + 1167.5 + 1785 = 5932.3$$

$$\text{Correction factor, CF} = G^2/n = (5932.3)^2/31 = 1135231.719$$

$$\text{SS due to treatments, SST} = \sum_{i=1}^v \frac{T_i^2}{r_i} - CF$$

$$= (562.4)^2/5 + (1261.6)^2/6 + (1155.8)^2/6 + (1167.5)^2/6 + (1785.0)^2/8 - 1135231.719$$

$$= 41399.233$$

$$\text{Total SS} = \sum_{i=1}^v \sum_{j=1}^{f_i} y_{ij}^2 - CF$$

$$= (108.2)^2 + (112.7)^2 + \dots + (227.6)^2 + (228.3)^2 - 1135231.719$$

$$= 49891.17$$

$$\text{Error SS} = \text{Total SS} - \text{SS due to treatments} = \text{SSE}$$

$$= 49891.17 - 41399.233 = 8491.938$$

The following Analysis of Variance Table is then formed.

**Table 2.3: ANOVA table for data in Example 1**

Source	DF	SS	MS	F-value	Prob > F
Treatments	4	41399.233	10349.808	31.69	<0.0001
Error	26	8491.938	326.613		
Total	30	49891.171			

R-square	CV	RMSE	Yield Mean
0.830	9.444	18.072	191.364

The model used has been able to explain 83 per cent of the total variability in the data. Since calculated  $F$  value = 31.69 is greater than the tabulated  $F$  at 4 and 26 degrees of freedom at 5% level of significance (= 2.742), the null hypothesis is rejected and at least two treatment effects are significantly different from each other at 5% level of significance. In fact, the probability of obtaining a value of  $F$  greater than 31.69 is smaller than 0.0001, meaning thereby that at least two treatments effects differ significantly even at smaller than 0.01% level of significance.

Now, to compare the treatment pairs, we calculate treatment means and LSD values at 5% level of significance. The treatment means are given in Table 2.4.

**Table 2.4: Treatment wise mean and standard deviation of dry matter accumulation**

Level of treatment	N	Dry matter accumulation	
		Mean	Standard Deviation
1	5	112.480	4.967
2	6	210.267	36.967
3	6	192.633	11.031
4	6	194.583	10.316
5	8	223.125	7.740



We now proceed to test the equality of treatments effects, *i.e.*  $H_0: \tau_i - \tau_l = 0$ , for all  $i \neq l = 1, 2, \dots, v$ . This is equivalent to making all the possible pairwise treatment comparisons. Table 2.5 gives  $|\text{difference between two treatment means}|$  and the least significant difference (LSD). Here  $|x|$  is the absolute value of  $x$ . In other words, it is the value of  $x$  ignoring the sign. If the difference of treatment means is larger than the LSD, then the two treatments are significantly different from each other at 5 per cent level of significance.

**Table 2.5: Least significant differences of treatment pairs**

Treatment Numbers	Difference of treatments means	Least Significant Difference
1, 2	97.79	22.35
1, 3	80.15	22.35
1, 4	82.1	22.35
1, 5	110.65	21.04
2, 3	17.64	21.31
2, 4	15.69	21.30
2, 5	12.86	19.93
3, 4	1.95	21.31
3, 5	30.5	19.93
4, 5	28.55	19.33

Alternatively, arrange the treatment means in ascending or descending order depending upon the character under study. If it is yield, it may be arranged in descending order and if it is disease infestation, it may be arranged in ascending order.

**Table 2.6: Treatment means arranged in descending order**

Dry matter accumulation	Treatment	Rank
223.125	5	1
210.267	2	2
194.583	4	3
192.633	3	4
112.480	1	5

Take the different between two treatment means with consecutive ranks. In Table 2.6, the difference  $\text{Trt}_5 - \text{Trt}_2 = 223.125 - 210.267 = 12.858$ . The LSD at 5% for these two treatments is 19.93. Therefore,  $\text{Trt}_5$  and  $\text{Trt}_2$  are not significantly different and may be assigned the same letter A. Since treatments ranked 1 and 2 are statistically not significant, therefore, now check the difference between treatment with rank 1 and rank 3, *i.e.*,  $\text{Trt}_5 - \text{Trt}_4 = 223.125 - 194.583 = 28.542$ . The LSD at 5% for these two treatments is 19.33. Therefore, these are statistically

different, or we can say that Treatment 5 is statistically better than treatment 3 and we may assign a different letter, say B, to treatment rank 3.

Since treatment with rank 1 is statistically significant compared to treatment with rank 3, now significance of treatment with rank 1 need not be tested with treatments with rank 4 and 5. Treatment with rank 1 will automatically be significantly different from treatments with ranks 4 and 5.

The procedure of making pairwise treatment comparisons just explained always holds when the estimated variance of the estimated difference of every possible pair of treatments is same as happens in CRD with equal replication, RCB design, LSD or any other variance balanced design. In case the estimated variance of the estimated difference of every possible pair of treatments is different, then we may need to check the significance of all treatment contrasts.

In this example, however, estimated variances of the estimated difference of all other pairs of treatments is less than that between treatments with rank 1 and rank 3. Therefore, we may stop checking significance of treatment with rank 1 with treatments with ranks 4 and 5. Now, start with treatment with rank 2 and test the significance of difference of treatment effects with rank 2 and 3 as  $\text{Trt}_2 - \text{Trt}_3 = 210.267 - 194.583 = 15.684$ , which is less than corresponding LSD at 5% level of significance (21.30). Therefore, we assign a second letter to treatment with rank 2 same as that was assigned to treatment with rank 3 earlier, i.e., B. Now, treatment with rank 2 has two symbols A and B depicting that it is not significantly different from treatments with rank 1 and 3. Now, we proceed to test the significance of difference of treatment effects with ranks 2 and 4 as  $\text{Trt}_2 - \text{Trt}_4 = 210.267 - 192.633 = 17.634$ , which is less than corresponding LSD at 5% level of significance (21.31). Therefore, now treatment 3 with rank 4 may also be assigned the same letter B. Next, proceed to test the significance of difference of treatment effects with ranks 2 and 5 as  $\text{Trt}_2 - \text{Trt}_5 = 210.267 - 112.480 = 97.787$ , which is more than corresponding LSD at 5% level of significance (22.35). Therefore, now treatment with rank 5 may be assigned a different letter, say C. Next, proceed to test significance of difference of effects of treatments with ranks 3 and 4, i.e.,  $\text{Trt}_4 - \text{Trt}_3 = 194.583 - 192.533 = 2.050$ , which is less than the LSD at 5%. Therefore, these two treatments are statistically at par and already assigned same letter B. We proceed with testing the same way and will get the Table 2.7.

**Table 2.7: Treatments with letter display**

Dry matter accumulation	Treatment	Rank
223.125A	5	1
210.267A,B	2	2
194.583B	4	3
192.633B	3	4
112.480C	1	5

Another way of presenting this Table is

Tr5	Tr2	Tr4	Tr3	Tr1
223.125	210.267	194.583	192.633	112.480

As per the Table 2.7, one can say that treatment 5 is significantly better than treatments 4, 3 and 1. Treatment 2 although statistically not significant with treatments 5, 4 and 3, is significantly different from treatment 1. Similarly treatments 4 and 3 are significantly different from treatment 1. Therefore, if treatment with highest mean is best, then any one of the treatments Tr5 or Tr2 may be used as they are statistically at par.

It may be noted that LSD controls only individual error rate and should be used only when null hypothesis of equality of treatment effects through ANOVA is rejected. Other commonly used multiple comparison procedure test that controls only individual error rate is Duncan's Multiple range Test. Some tests which control family error rate are Bonferroni correction and Tukey's Honestly Significant Differences (HSD) test and can be used even when null hypothesis through ANOVA is not rejected. More details on multiple comparison procedures may be seen in Annexure-III.

### 2.2.5 Analysis using SAS

The design is a CRD with  $v = 5$  treatments and  $n = 31$  observations. The data has been analyzed using SAS software. The commands and the data preparation are given in the sequel.

```
DATA crd;
INPUT trt dma;
/* trt denotes the treatment number and dma denotes the dry matter accumulation
in g/plant*/;
CARDS;
1      108.2
1      112.7
1      116.8
1      106.8
1      117.9
2      225.2
2      226.4
2      135.2
2      227.5
2      218.2
2      229.1
3      176.5
3      195.2
3      188.4
3      190.3
3      210.3
3      195.1
4      201.3
```

```

4      183.6
4      197.5
4      186.1
4      188.6
4      210.4
5      214.3
5      226.2
5      215.0
5      230.6
5      212.6
5      230.4
5      227.6
5      228.3
;
PROC GLM DATA=crd;
CLASS trt;
MODEL dma = trt;
MEANS trt;
LSMEANS trt/PDIFF LINES;
RUN;

```

**Remark 2.1** It may be worthwhile mentioning here that in the INPUT statement, CLASS statement and MODEL statement etc. the terms like trt, rep, etc. have been used to represent treatments, replications, etc. The output of analysis will also be using these notations. But while giving the results of analysis, the abbreviated forms are not used. Instead, the full forms are used for clarity and better understanding.

### 2.2.6 Output of analysis

The results obtained by the analysis using SAS are given in Table 2.8. This output is same as described earlier. The model with treatment effects only has been able to explain 83 per cent of the total variation. It is seen from the analysis of variance table that the treatment effects are significantly different ( $p$ -value < 0.0001).

**Table 2.8: SAS output for data in Example 1**

#### ANOVA

Source	SS	DF	MS	F-value	Prob > F
Model	41399.233	4	10349.808	31.69	<0.0001
Error	8491.938	26	326.613		
Total	49891.171	30			

R-square	CV	RMSE	dma Mean
0.830	9.444	18.072	191.364

## ANOVA

Source	Type I SS	DF	MS	F value	Prob > F
Treatments	41399.233	4	10349.808	31.69	<0.0001
Error	8491.938	26	326.613		
Total	49891.171	30			

The distribution of the observations for each treatment is given in the Figure 2.1.

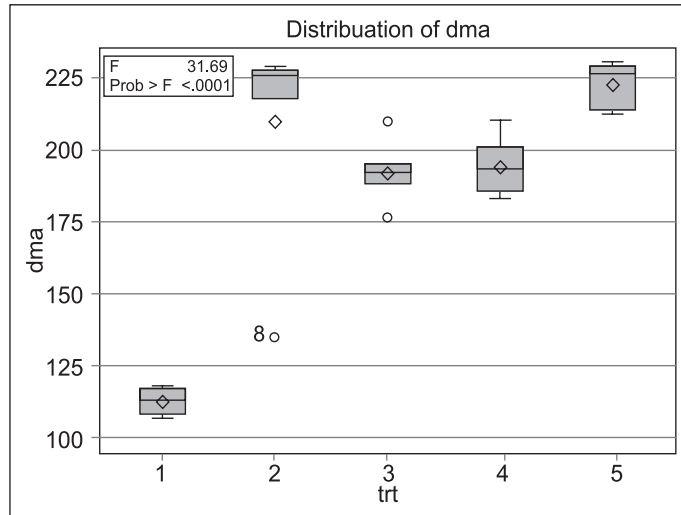


Figure 2.1: Treatment wise Box plot of dry matter accumulation

The mean and standard deviation of the dry matter accumulation for each of the treatments is given in Table 2.9.

**Table 2.9: Treatment wise mean and standard deviation of dry matter accumulation**

Level of treatment	N	Dry matter accumulation	
		Mean	Standard Deviation
1	5	112.480	4.967
2	6	210.267	36.967
3	6	192.633	11.031
4	6	194.583	10.317
5	8	223.125	7.7434

Table 2.10 gives the  $p$ -values for making pairwise treatment comparisons. These comparisons are similar to the one made above using LSD table. A  $p$ -value smaller than 0.05 implies that the pair of treatment effects is significantly different. For example, a  $p$ -value < 0.0001 indicates that

the treatment effects 1 and 2; 1 and 3; 1 and 4; 1 and 5; are significantly different. A *p*-value of 0.1030 suggests that the treatment effects 2 and 3 are not significantly different at 5 per cent level of significance. Similarly, a *p*-value of 0.1990 suggests that the treatment effects 2 and 5 are not significantly different at 5 per cent level of significance

**Table 2.10: P-values for pairwise comparison of the treatments**

Least Squares Means for effect treatment Pr >  t  for H0: LSMean(i)=LSMean(j) Dependent Variable: Dry Matter Accumulation					
i/j	1	2	3	4	5
1		<0.0001	<0.0001	<0.0001	<0.0001
2	<0.0001		0.1030	0.1450	0.1990
3	<0.0001	0.1030		0.8530	0.0040
4	<0.0001	0.1450	0.8530		0.0070
5	<0.0001	0.1990	0.0040	0.0070	

It may be noted from Table 2.11 that in this case the LS means are the same as the unadjusted means. Table 2.11 is another way of explaining the significance of difference of two treatment effects. Treatments with same letter are not significantly different.

**Table 2.11: Treatments with letter display**

<i>t</i> Comparison Lines for Least Squares Means of Treatments				
LS-means with the same letter are not significantly different				
		DMA LSMEAN	Treatment	Rank of Treatment
	A	223.125	5	5
B	A	210.267	2	4
B		194.583	4	3
B		192.633	3	2
	C	112.480	1	1

Since dry matter accumulation is highest for Trt5 and is significantly different from all other treatment effects, except treatment 2, so Trt5 *i.e.*, 50% recommended dose of fertilizers + vermicompost 2.5 tonnes/ha + farmyard manures 5 tonnes/ha is the best integrated nutrient management system, which is at par with 100% NPK (Trt2) so far as dry matter accumulation in tomato is concerned.

The pairwise treatment comparisons can also be made without writing the treatments in descending order of the treatments LS Mean values. The results are given in Table 2.12.

**Table 2.12: Treatments with letter display**

Treatment	DMA LS Mean	Rank of Treatment
1	112.480 <sup>C</sup>	5
2	210.267 <sup>A,B</sup>	2
3	192.633 <sup>B</sup>	4
4	194.583 <sup>B</sup>	3
5	233.125 <sup>A</sup>	1
General Mean	191.364	

In Table 2.12, any two treatments whose LS Means have at least one letter common are not statistically significant using LSD at given level of significance. Therefore, it follows that treatment 1 is significantly different from treatments 2, 3, 4, 5. Similarly, treatment 5 is significantly different from treatments 1, 3 and 4 but is not significantly different from treatment 2. On the other hand, treatment 2 is not significantly different from treatments 3, 4 and 5.

It may be worthwhile mentioning here that all comparisons are made at 5 per cent level of significance.

### 2.2.7 Analysis using R

The purpose of this section is to give the R code for analysis of data generated from a CRD for the benefit of the readers who would like to use R software. It may be mentioned here that the output obtained from R code is not given to avoid repetition.

```
d1=read.table("crd.txt",header=TRUE)
attach(d1)
names(d1)
#Treatment means and standard deviations
aggregate(dma, by=list(trt), mean)
aggregate(dma, by=list(trt), sd)
#Treatment wise box plot of dma
boxplot(dma~trt)
#ANOVA
trt=factor(trt)
crdout<-aov(dma~trt)
summary(crdout)
#Tukey's honest significant difference test is inbuilt part of Base R
TukeyHSD(crdout)
#LSD test, download and install agricolae package
library(agricolae)
lsd.result <- LSD.test(crdout,"trt")
lsd.result
detach(d1)
```

## 2.3 Randomized complete block design

A CRD assumes that there is no variability in the experimental units. The only source of variability in the data is the treatments. However, the experimental units selected for experimentation may exhibit variability because of many reasons. If the experimental units are the animals and the treatments are the grazing systems, then the initial body weight of the animal may be a major source of variability. Similarly, if the experimental units are plots in a field, and the treatments are the various levels of fertilizers and or irrigation, then the soil fertility may be a source of variability. Similarly with feeding trials in animal experiments, the lactation number may be a source of variability. Litter mates of animals may be source of variability in animal experiments. The salinity patches in the soil may be source of variability in field experiments. The variability in the experimental units needs to be accounted for, otherwise the experimental error will be unduly large and the Coefficient of variation (CV) would be overly large, which may lead to not rejecting the null hypothesis.

The focus of this Section is on designs useful for situations when there is heterogeneity in the experimental units and it is expected that there is only one source of variability in the experimental units. All the three principles of experimentation, viz., *randomization*, *replication* and *local control* are used in these designs. In these designs, the experimental units are partitioned into groups (called blocks) in such a way that experimental units within each block are as homogeneous as possible. As the name itself suggests, a Randomized Complete Block (RCB) design is a complete block design in the sense that each block is a complete replication. In other words, all the treatments in the experiment appear once in each block. Consequently, the block size, or the number of experimental units in each block is equal to the number of treatments. Further, since each block is a complete replication, the number of blocks is also equal to the replication number of treatments.

The randomization procedure in a RCB design is the following: (i) the treatments are randomly allocated the treatment labels, (ii) the treatments are assigned randomly to the experimental units within each block, and (iii) a separate randomization is done in each block.

The linear model in this case is

*Expected response = general mean + effect of treatments + effect due to experimental units (grouped as blocks)*

Since there is only one source of variation in the experimental units, the model can be rewritten as

*Expected response = general mean + effect of treatments + effect of blocks (or replications).*

This can also be written as

*response = general mean + treatments effect + block (or replication) effect + error,*

where the errors are independently distributed as normal variate with zero mean and constant variance  $\sigma^2$ . The split of the total variability in this case is



<b>Source of variation</b>
Due to model
Error
Total

The component “due to model” can be split as

<b>Source of variation</b>
Due to model
Due to treatments
Due to blocks (or replications)

### 2.3.1 Analysis of RCB design

Suppose that an experiment is run in a RCB design with  $v$  treatments and  $b$  replications (or complete blocks). Suppose that the observation generated on the response variable from the  $i$ th treatment in the  $j$ th block is represented by  $y_{ij}$ ,  $i = 1, 2, \dots, v$ ;  $j = 1, 2, \dots, b$ . The observations are represented by the following linear, additive model

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

where  $\mu$  is the general mean effect;  $\tau_i$  is the effect of the  $i$ th treatment (fixed);  $\beta_j$  is the effect of the  $j$ th block (fixed);  $e_{ij}$  is random error associated with  $y_{ij}$ , assumed to be mutually independent and distributed identically as normal variable with mean zero and common variance  $\sigma^2$ , i.e.,

$$E(e_{ij}) = 0, \quad \text{Cov}(e_{ij}, e_{kl}) = \begin{cases} \sigma^2, & \text{if } i = k, j = l \\ 0, & \text{otherwise} \end{cases}$$

Let the treatment totals and the block totals be denoted as respectively,  $T_i = \sum_{j=1}^b y_{ij}$ ,  $\forall i = 1, 2, \dots, v$

and  $B_j = \sum_{i=1}^v y_{ij}$ ,  $\forall j = 1, 2, \dots, b$ , and grand total as  $G = \sum_{i=1}^v T_i = \sum_{j=1}^b B_j = \sum_{i=1}^v \sum_{j=1}^b y_{ij}$ .

The following formulae can be employed for analysis of variance:

$$\text{Correction factor (CF)} = \frac{G^2}{vb}$$

$$\text{Total sum of squares} = \sum_{i=1}^v \sum_{j=1}^b y_{ij}^2 - CF$$

$$\text{Sum of squares due to treatments (SST)} = \sum_{i=1}^v \frac{T_i^2}{b} - CF$$

$$\text{Sum of squares due to blocks (or replications) (SSB)} = \sum_{i=1}^v \frac{T_i^2}{b} - CF$$

$$\text{Error sum of squares (SSE)} = \sum_{i=1}^v \sum_{j=1}^b y_{ij}^2 - \sum_{i=1}^v \frac{T_i^2}{b} - \sum_{j=1}^b \frac{B_j^2}{v} + CF$$

$$= \left( \sum_{i=1}^v \sum_{j=1}^b y_{ij}^2 - CF \right) - \left( \sum_{i=1}^v \frac{T_i^2}{b} - CF \right) - \left( \sum_{j=1}^b \frac{B_j^2}{v} - CF \right)$$

$$= \text{Total SS} - \text{Treatment SS} - \text{Block SS.}$$

The interest of the experimenter is in testing the null hypothesis:  $H_0 : \tau_1 = \tau_2 = \dots = \tau_v = \tau$  (say) against the alternative that  $\tau_i \neq \tau_l, i \neq l = 1, 2, \dots, v$  for at least one pair of treatment effects, say  $\tau_i$  and  $\tau_l$ . For testing this hypothesis we set up the analysis of variance Table 2.13.

**Table 2.13: ANOVA table for RCB design**

Source	DF	SS	MS = SS/DF	F
Treatments	$v - 1$	$SST = \sum_{i=1}^v \frac{T_i^2}{b} - CF$	$\frac{SST}{(v - 1)} = s_t^2$	$\frac{s_t^2}{s_e^2}$
Blocks (or Replications)	$b - 1$	$SSB = \sum_{j=1}^b \frac{B_j^2}{v} - CF$	$\frac{SSB}{(b - 1)} = s_b^2$	$\frac{s_b^2}{s_e^2}$
Error	$(v - 1)(b - 1)$	$SSE = \sum_{i=1}^v \sum_{j=1}^b y_{ij}^2 - \sum_{i=1}^v \frac{T_i^2}{b} - \sum_{j=1}^b \frac{B_j^2}{v} + CF$	$\frac{SSE}{(v - 1)(b - 1)} = s_e^2$	
Total	$vb - 1$	$\sum_{i=1}^v \sum_{j=1}^b y_{ij}^2 - CF$		

If the calculated value of  $F$  is greater than the table value of  $F_{\alpha, (v-1), (v-1)(b-1)}$  at  $\alpha$  level of significance and  $(v - 1), (v - 1)(b - 1)$  degrees of freedom, then the null hypothesis  $H_0$  is rejected at  $\alpha$  level of significance and it can be concluded that the treatment effects are significantly different from one another.

It may be seen here that the unbiased estimator of  $\sigma^2$  is  $\hat{\sigma}^2 = s_e^2$ .

Further, all the elementary treatment contrasts are estimable through the design. The Best Linear Unbiased Estimator (BLUE) of any treatment contrast  $d_{il} = \tau_i - \tau_l$  is

$$\hat{d}_{il} = \frac{T_i - T_l}{b}, \forall i \neq l = 1, 2, \dots, v.$$

The variance of  $\hat{d}_{il}$  is  $\text{Var}(\hat{d}_{il}) = \frac{2\sigma^2}{b}$ . The estimated standard error of the estimated difference

between the  $i$ th and  $l$ th treatment effects is  $SE(\hat{d}_{il}) = \left( \frac{2s_e^2}{b} \right)^{1/2}$ .

The Least Significant Difference (LSD) at  $\alpha$  level of significance is given as

$$\text{LSD} = SE(\hat{d}_{il}) \times t_{\alpha/2; \text{errorDF}}$$

Here  $t_{\alpha/2; \text{errorDF}}$  denotes the value of Student's  $t$  at  $\alpha$  level of significance and error degrees of freedom. The treatment means are given by  $\bar{T}_i = \frac{T_i}{b} \forall i = 1, 2, \dots, v$ . The pairwise comparison of treatment effects can be made by comparing the difference between any two treatment means with the *LSD*. Any two treatment effects are said to differ significantly if the difference of their means is larger than the *LSD*.

### 2.3.2 Example 2

An initial varietal trial (Late Sown, irrigated) was conducted to study the performance of 20 new strains of mustard vis-a-vis four checks (Swarna Jyoti: ZC; Vardan: NC; Varuna: NC; and Kranti: NC) using a Randomized Complete Block Design (RCB) design at Bhatinda with 3 replications under the aegis of All India Coordinated Research Project on Rapeseed and Mustard. The seed yield in kg/ha was recorded. The details of the experiment are given in Table 2.14.

In the sequel, the data are analyzed (a) to test whether or not there is any difference among the treatment effects, (b) to make all the possible pairwise treatment comparisons to identify the best treatment *i.e.* the treatment giving highest yield, and (c) to test whether or not the average performance of check varieties (i) Swarna Jyoti (MCN-04-128), (ii) Vardan (MCN-04-129), (iii) Varuna (MCN-04-131), and (iv) Kranti (MCN-04-133) is significantly different from average performance of remaining strains.

**Table 2.14: Seed yield (kg/ha) data**

Treatment Number	Strain	Code	Replications		
			1	2	3
1	RK-04-3	MCN-04-110	1539.69	1412.35	1319.73
2	RK-04-4	MCN-04-111	1261.85	1065.05	1111.36
3	RGN-124	MCN-04-112	1389.19	1516.54	1203.97
4	HYT-27	MCN-04-113	1192.39	1215.55	1157.66
5	PBR-275	MCN-04-114	1250.27	1203.97	1366.04
6	HUJM-03-03	MCN-04-115	1296.58	1273.43	1308.16
7	RGN-123	MCN-04-116	1227.12	1018.74	937.71
8	BIO-13-01	MCN-04-117	1273.43	1157.66	1088.20
9	RH-0115	MCN-04-118	1180.82	1203.97	1041.90
10	RH-0213	MCN-04-119	1296.58	1458.65	1250.27
11	NRCDR-05	MCN-04-120	1122.93	1065.05	1018.74
12	NRC-323-1	MCN-04-121	1250.27	926.13	1030.32
13	RRN-596	MCN-04-122	1180.82	1053.47	717.75
14	RRN-597	MCN-04-123	1146.09	1180.82	856.67
15	CS-234-2	MCN-04-124	1574.42	1412.35	1597.57
16	RM-109	MCN-04-125	914.55	972.44	659.87
17	BAUSM-2000	MCN-04-126	891.40	937.71	798.79
18	NPJ-99	MCN-04-127	1227.12	1203.97	1389.19
19	<b>SWARNA JYOTI(ZC)</b>	<b>MCN-04-128</b>	1389.19	1180.82	1273.43
20	<b>VARDAN(NC)</b>	<b>MCN-04-129</b>	1331.31	1157.66	1180.82
21	PR-2003-27	MCN-04-130	1250.27	1250.27	1296.58
22	<b>VARUNA(NC)</b>	<b>MCN-04-131</b>	717.75	740.90	578.83
23	PR-2003-30	MCN-04-132	1169.24	1157.66	1111.36
24	<b>KRANTI-(NC)</b>	<b>MCN-04-133</b>	1203.97	1296.58	1250.27

**Note:** Strains of mustard in bold are the four checks.

### 2.3.3 Analysis of data

#### Treatment Totals

$$T_1 = 1539.69 + 1412.35 + 1319.73 = 4271.77$$

$$T_2 = 1261.85 + 1065.05 + 1111.36 = 3438.26$$

$$T_3 = 1389.19 + 1516.54 + 1203.97 = 4109.70$$

$$T_4 = 1192.39 + 1215.55 + 1157.66 = 3565.60$$

$$T_5 = 1250.27 + 1203.97 + 1366.04 = 3820.28$$

$$T6 = 1296.58 + 1273.43 + 1308.16 = 3878.17$$

$$T7 = 1227.12 + 1018.74 + 937.71 = 3183.57$$

$$T8 = 1273.43 + 1157.66 + 1088.20 = 3519.29$$

$$T9 = 1180.82 + 1203.97 + 1041.90 = 3426.69$$

$$T10 = 1296.58 + 1458.65 + 1250.27 = 4005.50$$

$$T11 = 1122.93 + 1065.05 + 1018.74 = 3206.72$$

$$T12 = 1250.27 + 926.13 + 1030.32 = 3206.72$$

$$T13 = 1180.82 + 1053.47 + 717.75 = 2952.04$$

$$T14 = 1146.09 + 1180.82 + 856.67 = 3183.58$$

$$T15 = 1574.42 + 1412.35 + 1597.57 = 4584.34$$

$$T16 = 914.55 + 972.44 + 659.87 = 2546.86$$

$$T17 = 891.40 + 937.71 + 798.79 = 2627.90$$

$$T18 = 1227.12 + 1203.97 + 1389.19 = 3820.28$$

$$T19 = 1389.19 + 1180.82 + 1273.43 = 3843.44$$

$$T20 = 1331.31 + 1157.66 + 1180.82 = 3669.79$$

$$T21 = 1250.27 + 1250.27 + 1296.58 = 3797.12$$

$$T22 = 717.75 + 740.90 + 578.83 = 2037.48$$

$$T23 = 1169.24 + 1157.66 + 1111.36 = 3438.26$$

$$T24 = 1203.97 + 1296.58 + 1250.27 = 3750.82$$

### Block Totals

$$B1 = 1539.69 + 1261.85 + 1389.19 + \dots + 717.75 + 1169.24 + 1203.97 = 29277.25$$

$$B2 = 1412.35 + 1065.05 + 1516.54 + \dots + 740.90 + 1157.66 + 1296.58 = 28061.74$$

$$B3 = 1319.73 + 1111.36 + 1203.97 + \dots + 578.83 + 1111.36 + 1250.27 = 26545.19$$

$$\text{Grand Total, } G = 29277.25 + 28061.74 + 26545.19 = 4271.77 + 3438.26 + 4109.70 + \dots + 2037.48 + 3438.26 + 3750.82 = 83884.18$$

$$\text{Correction Factor, } CF = \frac{(83884.18)^2}{72} = \frac{7036555654}{72} = 97729939.64$$

$$\text{Treatments SS (SST)} = \frac{(4271.77)^2 + (3438.26)^2 + \dots + (3438.26)^2 + (3750.82)^2}{3} - CF$$

$$= 100244098.93 - 97729939.64 = 2514159.29$$

$$\text{Block (or Replication) SS (SSB)} = \frac{(29277.25)^2 + (28061.74)^2 + (26545.19)^2}{24} - CF$$

$$= 97886072.15 - 97729939.64 = 156132.51$$

$$\text{Total SS} = (1539.69)^2 + (1412.35)^2 + (1319.73)^2 + (1261.85)^2 + (1065.05)^2 + \dots + (1203.97)^2 + (1269.58)^2 + (1250.27)^2 - CF$$

$$= 100863347.59 - 97729939.64 = 3133407.95$$

$$\text{Error SS (SSE)} = \text{Total SS} - \text{Treatments SS} - \text{Replication (or Block) SS}$$

$$= 3133407.95 - 2514159.29 - 156132.51 = 463116.15$$

We then have the analysis of variance as shown in Table 2.15.

**Table 2.15: ANOVA table for the data in Example 2**

Source	DF	SS	MS	F-value	Prob > F
Treatments	23	2514159.289	109311.273	10.86	<0.0001
Blocks (or Replications)	2	156132.504	78066.250	7.75	0.0013
Error	46	463116.156	10067.743		
Total	71	3133407.949			

This analysis reveals that the treatment differences are highly significant ( $p$ -value < 0.0001). Similarly, the block effects are also highly significant ( $p$ -value = 0.0013) meaning thereby that the block formation has proved to be very effective. The blocks formation was genuinely required and blocks formation has been proper.

### 2.3.4 Analysis using SAS

The design is a RCB design with  $v = 24$  treatments,  $b = 3$  blocks (or replications) and  $n = 72$  observations. The data has been analyzed using SAS software. The commands and the data preparation are given in the sequel.

```
DATA rbd; /*one can enter any other name for Data*/;
INPUT trt $ 11. trtn rep syield;
*here 11. represents that the value of the variable trt is upto 11 columns;
/*trtn denotes the treatment number, rep the replication number and syield
the seed yield in kg/hectare*/;
CARDS;
MCN-04-110          1          1          1539.69
MCN-04-111          2          1          1261.85
MCN-04-112          3          1          1389.19
```

MCN-04-113	4	1	1192.39
MCN-04-114	5	1	1250.27
MCN-04-115	6	1	1296.58
MCN-04-116	7	1	1227.12
MCN-04-117	8	1	1273.43
MCN-04-118	9	1	1180.82
MCN-04-119	10	1	1296.58
MCN-04-120	11	1	1122.93
MCN-04-121	12	1	1250.27
MCN-04-122	13	1	1180.82
MCN-04-123	14	1	1146.09
MCN-04-124	15	1	1574.42
MCN-04-125	16	1	914.55
MCN-04-126	17	1	891.40
MCN-04-127	18	1	1227.12
MCN-04-128	19	1	1389.19
MCN-04-129	20	1	1331.31
MCN-04-130	21	1	1250.27
MCN-04-131	22	1	717.75
MCN-04-132	23	1	1169.24
MCN-04-133	24	1	1203.97
MCN-04-110	1	2	1412.35
MCN-04-111	2	2	1065.05
MCN-04-112	3	2	1516.54
MCN-04-113	4	2	1215.55
MCN-04-114	5	2	1203.97
MCN-04-115	6	2	1273.43
MCN-04-116	7	2	1018.74
MCN-04-117	8	2	1157.66
MCN-04-118	9	2	1203.97
MCN-04-119	10	2	1458.65
MCN-04-120	11	2	1065.05
MCN-04-121	12	2	926.13
MCN-04-122	13	2	1053.47
MCN-04-123	14	2	1180.82
MCN-04-124	15	2	1412.35
MCN-04-125	16	2	972.44
MCN-04-126	17	2	937.71
MCN-04-127	18	2	1203.97
MCN-04-128	19	2	1180.82
MCN-04-129	20	2	1157.66
MCN-04-130	21	2	1250.27
MCN-04-131	22	2	740.90

MCN-04-132	23	2	1157.66
MCN-04-133	24	2	1296.58
MCN-04-110	1	3	1319.73
MCN-04-111	2	3	1111.36
MCN-04-112	3	3	1203.97
MCN-04-113	4	3	1157.66
MCN-04-114	5	3	1366.04
MCN-04-115	6	3	1308.16
MCN-04-116	7	3	937.71
MCN-04-117	8	3	1088.20
MCN-04-118	9	3	1041.90
MCN-04-119	10	3	1250.27
MCN-04-120	11	3	1018.74
MCN-04-121	12	3	1030.32
MCN-04-122	13	3	717.75
MCN-04-123	14	3	856.67
MCN-04-124	15	3	1597.57
MCN-04-125	16	3	659.87
MCN-04-126	17	3	798.79
MCN-04-127	18	3	1389.19
MCN-04-128	19	3	1273.43
MCN-04-129	20	3	1180.82
MCN-04-130	21	3	1296.58
MCN-04-131	22	3	578.83
MCN-04-132	23	3	1111.36
MCN-04-133	24	3	1250.27

;

RUN;

PROC GLM ;

CLASS trtn rep;

MODEL syield = trtn rep;

LSMEANS trtn/PDIFF LINES;

CONTRAST 'check vs strains' trtn 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 -20 -20 4 -20 4 -20;

RUN;

In order to compare the check varieties with the strains, the null hypothesis to be tested is that the average effect of strains is same as the average effect of check varieties. The null hypothesis  $H_0: 4\tau_1 + 4\tau_2 + \dots + 4\tau_{17} + 4\tau_{18} + 4\tau_{21} + 4\tau_{23} - 20\tau_{19} - 20\tau_{20} - 20\tau_{22} - 20\tau_{24} = 0$  is tested against

$$H_1: 4\tau_1 + 4\tau_2 + \dots + 4\tau_{17} + 4\tau_{18} + 4\tau_{21} + 4\tau_{23} - 20\tau_{19} - 20\tau_{20} - 20\tau_{22} - 20\tau_{24} \neq 0.$$



For testing the hypothesis  $H_0$ , one needs to perform contrast analysis. The problem of contrast analysis has been dealt with in Chapter 3. However, here we have given the SAS steps to perform the contrast analysis.

The output of analysis using SAS is given in Table 2.16.

**Table 2.16: Output of analysis using SAS**

**ANOVA**

Source	DF	SS	MS	F-value	Prob > F
Model	25	2670291.793	106811.672	10.61	<0.0001
Error	46	463116.156	10067.743		
Corrected Total	71	3133407.949			

R-square	CV	Root MSE	Yield Mean
0.852	8.612	100.34	1165.06

**ANOVA**

Source	DF	SS	MS	F-value	Prob > F
Treatments	23	2514159.289	109311.273	10.86	<0.0001
Blocks	2	156132.504	78066.252	7.75	0.0013
Error	46	463116.156	10067.743		
Corrected Total	71	3133407.949			

The model with treatment effects and block effects explains about 85 per cent of the total variability in the data. The treatment effects are highly significant ( $p$ -value < 0.0001) meaning thereby that the null hypothesis is rejected. It is interesting to note that the block effects are also highly significant ( $p$ -value = 0.0013).

The mean and standard deviation of the treatments are given in Table 2.17.

**Table 2.17: Treatment wise mean and standard deviation of seed yield**

Level of treatment	N	SYIELD	
		Mean	Standard Deviation
1	3	1423.92	110.44
2	3	1146.09	102.89
3	3	1369.90	157.18
4	3	1188.53	29.14
5	3	1273.43	83.48
6	3	1292.72	17.68
7	3	1061.19	149.30

8	3	1173.10	93.57
9	3	1142.23	87.66
10	3	1335.17	109.42
11	3	1068.91	52.20
12	3	1068.91	165.48
13	3	984.01	239.22
14	3	1061.19	177.97
15	3	1528.11	100.92
16	3	848.95	166.29
17	3	875.97	70.73
18	3	1273.43	100.92
19	3	1281.15	26.74
20	3	1223.26	30.63
21	3	1265.71	104.40
22	3	679.16	94.28
23	3	1146.09	87.66
24	3	1250.27	46.31

The distribution of observations over replications for each treatment is given in Figure 2.2.

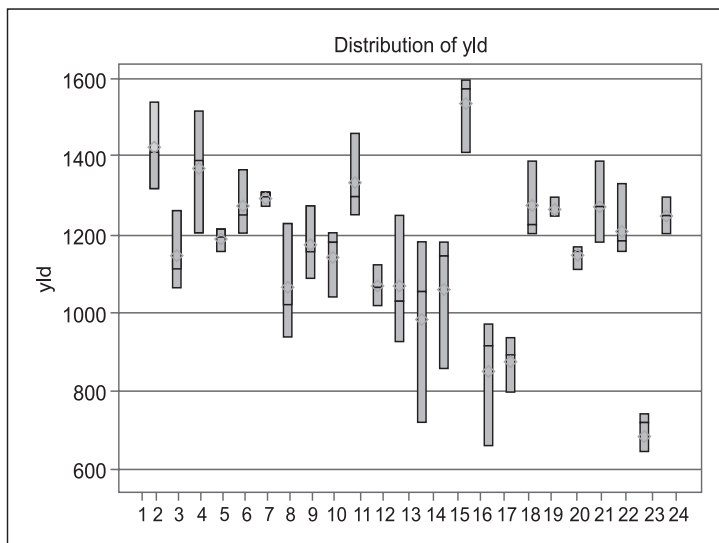


Figure 2.2: Treatment wise Box plot of seed yield

Similarly, the Figure 2.3 gives the plot of observations in each block. It is quite evident from

here that the blocks differ in their effects, *i.e.* mean square between blocks is high as compared to mean square error, a fact supported by the ANOVA as well.

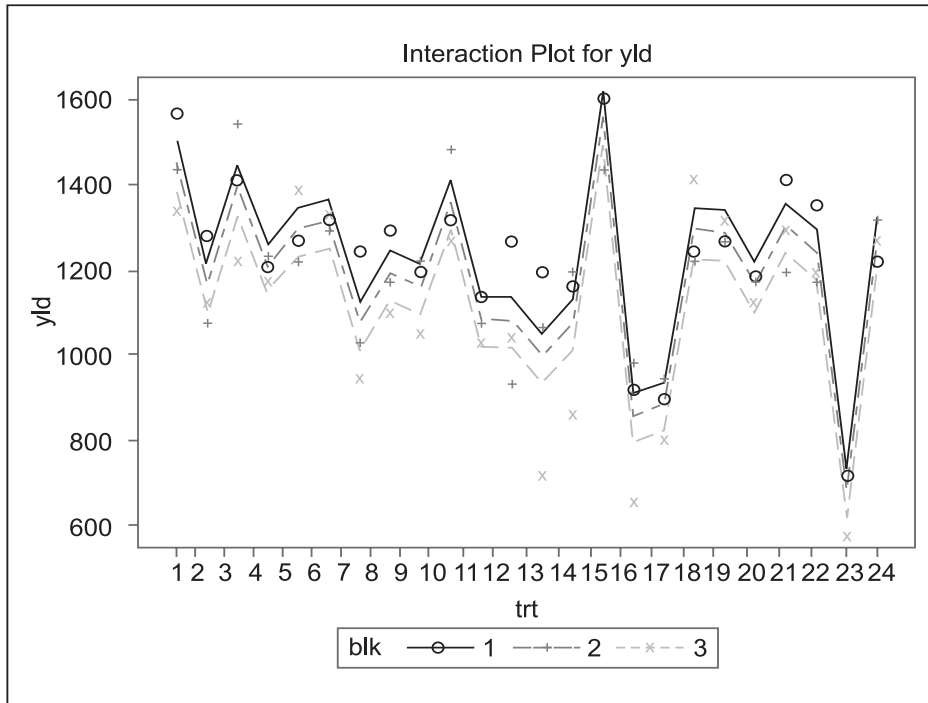


Figure 2.3: Plot of observations in each block

The pairwise comparison of treatment effects is made and is presented in Table 2.18. Treatments having at least one letter common are not significantly different in their effects. The strain CS-234-2 and coded as MCN-04-124 (treatment 15) is the highest seed yielding strain. This strain produces significantly higher seed yield than all other treatments produce except treatment numbers 1 and 3, which produce seed yield statistically at par with the produce of treatment 15. So this strain may be recommended as the best among the lot.

**Table 2.18: Treatments in descending order with letter display**

Means with the same letter are not significantly different							
Duncan Grouping					Mean	Treatment Number	Replication
A					1528.11	15	3
A	B				1423.92	1	3
A	B	C			1369.90	3	3
	B	C	D		1335.17	10	3
	B	C	D		1292.72	6	3
	B	C	D		1281.15	19	3
	B	C	D		1273.43	18	3
	B	C	D		1273.43	5	3
	B	C	D		1265.71	21	3
	B	C	D	E	1250.27	24	3
		C	D	E	1223.26	20	3
		C	D	E	1188.53	4	3
F			D	E	1173.10	8	3
F			D	E	1146.09	23	3
F			D	E	1146.09	2	3
F			D	E	1142.23	9	3
F				E	1068.91	11	3
F				E	1068.91	12	3
F				E	1061.19	14	3
F				E	1061.19	7	3
F	G				984.01	13	3
	G				875.97	17	3
	G				848.95	16	3
		H			679.16	22	3

From Table 2.18 it is also evident that check variety (Treatment 19: best performing check) is significantly different from strains at Treatment 7, 11, 12, 13, 14, 15, 16, 17. Similarly, check variety (Treatment 20) is significantly different from strains at Treatment 1, 13, 15, 16, 17. Further, check variety (Treatment 24) is significantly different from strains at Treatment 13, 15, 16, 17. The check variety (Treatment 22) is, however, the lowest yielding and is significantly different from all the strains.

The contrast analysis for testing the null hypothesis that the average effect of strains is same as the average effect of check varieties was done and the result is given in Table 2.19.

**Table 2.19: Result of contrast analysis**

Contrast	DF	Type III SS	MS	F-value	Prob > F
Check vs Strains	1	46126.736	46126.736	4.58	0.0377

It may be noted that the check varieties differ significantly from the strains ( $p$ -value = 0.0377).

The pairwise treatment comparisons can also be presented without writing the treatments in descending order of the treatments LS Mean values. The results are given in Table 2.20.

**Table 2.20: Treatments with letter display**

Treatment Name	LS Mean of Syield	Rank of Treatment	Treatment Name	LS Mean of Syield	Rank of Treatment
1	1423.92 <sup>A,B</sup>	2	13	984.01 <sup>G,H</sup>	21
2	1146.09 <sup>E,F,G</sup>	15	14	1061.19 <sup>FG</sup>	19
3	1369.90 <sup>A,B,C</sup>	3	15	1528.11 <sup>A</sup>	1
4	1188.53 <sup>D,E,F</sup>	12	16	848.95 <sup>H</sup>	23
5	1273.43 <sup>B,C,D,E</sup>	7	17	875.97 <sup>H</sup>	22
6	1292.72 <sup>B,C,D,E</sup>	5	18	1273.43 <sup>B,C,D,E</sup>	8
7	1061.19 <sup>FG</sup>	20	19	1281.15 <sup>B,C,D,E</sup>	6
8	1173.10 <sup>D,E,F</sup>	13	20	1223.26 <sup>C,D,E,F</sup>	11
9	1142.23 <sup>E,F,G</sup>	16	21	1265.71 <sup>B,C,D,E</sup>	9
10	1335.17 <sup>B,C,D</sup>	4	22	679.16 <sup>I</sup>	24
11	1068.91 <sup>FG</sup>	17	23	1146.09 <sup>E,F,G</sup>	14
12	1068.91 <sup>FG</sup>	18	24	1250.27 <sup>C,D,E</sup>	10
General Mean			1165.06		
$SE(\hat{d})$			81.927		
LSD at 5%			164.91		

It may be mentioned here that the SAS commands given in Section 2.3.4 do not compute LSD at 5%. This may, therefore, be computed by using the formula given in Section 2.3.1. One may also compute it by adding a SAS command “MEANS trtn/LSD;”. In Table 2.20, any two treatments whose LS Means have at least one letter common are not statistically significant using LSD. Therefore, it follows that treatment 15 is the one that produces highest seed yield and is not significantly different from treatments 1 and 3. It is significantly different from all the remaining treatments. Similarly, treatment 22, a control variety, produces the lowest seed yield and has in fact statistically significant lower seed yield from all other strains. The other three control varieties (Treatments 19, 20 and 24) are statistically at par with each other.

### 2.3.5 Analysis of RCB design using R

R code

```
d2=read.table("rbd.txt",header=TRUE)
attach(d2)
names(d2)
#Treatment means and standard deviations
aggregate(syield, by=list(trt), mean)
aggregate(syield, by=list(trt), sd)
#Treatment wise box plot of yield
boxplot(syield~trtn)
#ANOVA
trtn=factor(trtn)
rep=factor(rep)
aov.out=aov(syield~trtn+rep)
summary(aov.out)
library(lsmmeans)
lsm <- lsmmeans(aov.out, "trtn")
contrast(lsm, list(con1 = c(4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,-20,-20,4,-20,4,-20)))
#Tukey's honest significant difference test
TukeyHSD(rbdout)
#LSD
library(agricolae)
lsd.result <- LSD.test(aov.out,"trtn")
lsd.result
detach(d2)
```

## 2.4 Latin square design

A CRD assumes that there is no variability in the experimental units. The only source of variability in the data is the treatments and the remaining variability is the error. On the other hand, an RCB design assumes that other than the treatments, there is one source of variability in the experimental units and this variability in the experimental units is controlled by forming blocks of homogeneous experimental units. In this case, the sources of variability in the data are the treatments and the blocks (or replications) and the remaining part of the variability is the experimental error. This section is devoted to designs which control two sources of variability in the experimental units. When there are two sources of variability in the experimental units, we need to form blocks in two directions, perpendicular to each other. The two blocking systems are cross classified as rows and columns and the intersection of rows and columns is a cell or the experimental unit. Following on the example of four grazing systems and 16 experimental units (animals), one source of variability in the animals could be the initial body weight. The other source of variability could be their physiological behavior. For instance, the calving age or the number of lactations could be another source of variability in the experimental material. The physiological behavior and the initial body weights are the two sources of variability in

the animals and need to be controlled by proper designing of experiment. In the insecticide field trial where the insect migration has a predictable direction that is perpendicular to the dominant fertility gradient of the experimental field, there are two sources of variability in the experimental units. In order to control two-way heterogeneity in the experimental material, we use designs known as *Latin Square* obtained from Latin square arrangement. The following are examples of  $4 \times 4$  and  $5 \times 5$  Latin square designs:

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>		1	2	3	4
<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>		<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>		3	4	1	2
<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>	;	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>		4	3	2	1
<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>		<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>		2	1	4	3
					<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>					

In such designs two restrictions are imposed by forming blocks in two directions, row-wise and column-wise. A Latin square arrangement is an arrangement of  $\nu$  Latin letters in a  $\nu \times \nu$  square in such a way that each row and each column has all the  $\nu$  Latin letters appearing exactly once. A design based upon a Latin square arrangement is called a Latin square design. Ignoring rows and considering columns as blocks gives an RCB design. Similarly, ignoring columns and treating rows as blocks gives an RCB design. So a Latin square design is an RCB design in rows as well as columns. Treatments are allocated in such a way that every treatment occurs once and only once in each row and each column. In this design, the replication number of treatments is same as the number of treatments.

Latin squares have been classified as reduced and standard. The Latin squares have also been classified as squares with normalized or standard and semi-standard form, whereby reduced Latin square is synonym to normalized or standard form and standard Latin square is synonym to semi-standard form. Latin square is considered reduced if its first row and first column contains elements in the numerical  $(1, 2, \dots, \nu)$  or lexicographic order  $(A, B, C, \dots)$ . On the other hand, it is considered standard if only its first row contains elements in the natural order. In the examples of Latin squares given earlier, the first two Latin squares of order 4 and 5, respectively are the reduced squares (or Latin squares in standard form) while the third Latin square of order 4 is in standard (or semi-standard) form.

The randomization of the  $\nu$  treatments over the  $\nu^2$  experimental units arranged in a  $\nu \times \nu$  square is difficult. The design obtained after randomization should be a Latin square design. In actual field arrangement during experimentation, first we select a  $\nu \times \nu$  reduced or normalized or standard Latin square randomly from the Fisher and Yates Tables. Having selected the square, column-wise randomization is carried out first, followed by row-wise randomization. Of course, the treatments labels (or the Latin letters) are randomized separately before starting the actual randomization in the design.

The linear model in this case is

*Expected response = general mean + effect of treatments + effect due to experimental units*

Since there are two sources of variation in the experimental material, the model can be rewritten as

*Expected response = general mean + effect of treatments+ effect of rows + effect of columns*

This can also be written as:

*Response = general mean + treatments effect + rows effect + columns effect + error,*

where the errors are distributed independently as normal variate with zero mean and constant variance  $\sigma^2$ . The partitioning of the total variability in this case is

Source of Variation
Due to model
Error
Total

The component “due to model” can be split as

Source of Variation
Due to model
Due to Treatments
Due to Rows
Due to Columns

### 2.4.1 Analysis of Latin square design

The  $v^2$  observations generated from a Latin square design of order  $v$  are represented by the following linear, additive, fixed effects model:

$$y_{i(j,k)} = \mu + \tau_i + \beta_j + \gamma_k + e_{i(j,k)}; \quad i = 1, 2, \dots, v; j = 1, 2, \dots, v; k = 1, 2, \dots, v,$$

where  $y_{i(j,k)}$  is the observation pertaining to the  $i$ th treatment appearing in the  $(j, k)$ th cell,  $\mu$  is the grand mean,  $\tau_i$  is the  $i$ th treatment effect,  $\beta_j$  is the effect of the  $j$ th row,  $\gamma_k$  is the effect of the  $k$ th column, and  $e_{i(j,k)}$  is the random error associated with  $y_{i(j,k)}$ , assumed to be mutually independent and distributed normally with mean zero and common variance  $\sigma^2$ .

Let the treatment totals, rows totals, column totals be denoted as, respectively,

$$T_i = \sum_{(j,k) \supseteq i} y_{i(j,k)}, \quad \forall i = 1, 2, \dots, v; \text{ sum of observations over cells containing treatment } i;$$

$$R_j = \sum_{i=1}^v \sum_{k=1}^v y_{i(j,k)}, \quad \forall j = 1, 2, \dots, v; \text{ sum of observations in the } j\text{th row};$$



$C_k = \sum_{i=1}^v \sum_{j=1}^v y_{i(j,k)}$ ,  $\forall k = 1, 2, \dots, v$ ; sum of observations in the  $k$ th column.

The grand total is,  $G = \sum_{i=1}^v T_i = \sum_{j=1}^v R_j = \sum_{k=1}^v C_k = \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}$

The following formulae can be employed for analysis of variance:

$$\text{Correction factor (CF)} = \frac{G^2}{v^2}$$

$$\text{Total sum of squares (SS)} = \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}^2 - CF$$

$$\text{Sum of Squares due to treatments (SST)} = \frac{1}{v} \sum_{i=1}^v T_i^2 - CF$$

$$\text{Sum of Squares due to rows (SSR)} = \frac{1}{v} \sum_{j=1}^v R_j^2 - CF$$

$$\text{Sum of Squares due to columns (SSC)} = \frac{1}{v} \sum_{k=1}^v C_k^2 - CF$$

Error sum of squares (SSE)

$$= \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}^2 - \frac{1}{v} \sum_{i=1}^v T_i^2 - \frac{1}{v} \sum_{j=1}^v R_j^2 - \sum_{k=1}^v C_k^2 + 2(CF)$$

$$= \left( \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}^2 - CF \right) - \left( \frac{1}{v} \sum_{i=1}^v T_i^2 - CF \right) - \left( \frac{1}{v} \sum_{j=1}^v R_j^2 - CF \right) - \left( \frac{1}{v} \sum_{k=1}^v C_k^2 - CF \right)$$

= Total SS – Row SS – Column SS – Treatment SS

If the calculated value of  $F$  is greater than the table value of  $F_{\alpha; (v-1), (v-1)(v-2)}$  at  $\alpha$  level of significance and  $(v-1), (v-1)(v-2)$  degrees of freedom, then the null hypothesis  $H_0$  is rejected at  $\alpha$  level of significance and it can be concluded that the treatment effects are significantly different from one another.

It may be seen here that an unbiased estimator of  $\sigma^2$  is  $\hat{\sigma}^2 = s_e^2$ .

Further, all the elementary treatment contrasts are estimable through the design. The Best Linear Unbiased Estimator (BLUE) of any treatment contrast  $d_{il} = \tau_i - \tau_l$  is

$$\hat{d}_{il} = \frac{T_i - T_l}{v}, \forall i \neq l = 1, 2, \dots, v.$$

**Table 2.21: ANOVA table in Latin square design**

Source	DF	SS	MS	F
Treatments	$v - 1$	$SST = \frac{1}{v} \sum_{i=1}^v T_i^2 - CF$	$\frac{SST}{(v-1)} = s_t^2$	$\frac{s_t^2}{s_e^2}$
Rows	$v - 1$	$SSR = \frac{1}{v} \sum_{j=1}^v R_j^2 - CF$	$\frac{SSR}{(v-1)} = s_r^2$	$\frac{s_r^2}{s_e^2}$
Columns	$v - 1$	$SSC = \frac{1}{v} \sum_{k=1}^v C_k^2 - CF$	$\frac{SSC}{(v-1)} = s_c^2$	$\frac{s_c^2}{s_e^2}$
Error	$(v-1)(v-2)$	$SSE = \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}^2 - \frac{1}{v} \sum_{i=1}^v T_i^2 - \frac{1}{v} \sum_{j=1}^v R_j^2 - \sum_{k=1}^v C_k^2 + 2(CF)$	$\frac{SSE}{(v-1)(v-2)} = s_e^2$	
Total	$v^2 - 1$	$\sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{i(j,k)}^2 - CF$		

The variance of  $\hat{d}_{il}$  is  $\text{Var}(\hat{d}_{il}) = \frac{2\sigma^2}{v}$ . The estimated standard error of the difference between the

estimated  $i$ th and  $l$ th treatment effects is  $SE(\hat{d}_{il}) = \left( \frac{2s_e^2}{v} \right)^{1/2}$ .

The Least Significant Difference ( $LSD$ ) is given as  $LSD = SE(\hat{d}_{il}) \times t_{\alpha; error DF}$ .

Here  $t_{\alpha/2; error DF}$  denotes the value of Student's  $t$  at  $\alpha/2$  level of significance and error degrees of freedom. The treatment means are given by  $\bar{T}_i = \frac{T_i}{v} \forall i = 1, 2, \dots, v$ . The pairwise comparison of

treatment effects can be made by comparing the difference between any two treatment means with the  $LSD$ . Any two treatment effects are said to differ significantly if the difference of their means is larger than the  $LSD$ .

### 2.4.2 Example 3

An experiment was conducted at Agricultural Research Station, Kopurgaon, Maharashtra on Cotton using a Latin Square Design to study the effects of foliar application of urea in combination with insecticidal sprays on the cotton yield. The 6 treatments were  $\{T_1$ : Control (*i.e.* no  $N$  and no insecticides),  $T_2$ : 100kg  $N$ /ha applied as urea (half at final thinning and half at flowering as top dressing),  $T_3$ : 100kg  $N$ /ha applied as urea (80 kg  $N$ /ha in 4 equal split doses as spray and 20 kg  $N$ /ha at final thinning),  $T_4$ : 100 kg  $N$ /ha applied as CAN (half at final thinning and half at flowering as top dressing),  $T_5$ :  $T_2$ + six insecticidal sprays,  $T_6$ :  $T_4$  + six insecticidal

sprays}. There were 6 replications, and the data of Cotton yield in kg per plot is given in Table 2.22.

**Table 2.22: Cotton yield data (kg/plot)**

$T_3$ 3.10	$T_6$ 5.95	$T_1$ 1.75	$T_5$ 6.40	$T_2$ 3.85	$T_4$ 5.30
$T_2$ 4.80	$T_1$ 2.70	$T_3$ 3.30	$T_6$ 5.95	$T_4$ 3.70	$T_5$ 5.40
$T_1$ 3.00	$T_2$ 2.95	$T_5$ 6.70	$T_4$ 5.95	$T_6$ 7.75	$T_3$ 7.10
$T_5$ 6.40	$T_4$ 5.80	$T_2$ 3.80	$T_3$ 6.55	$T_1$ 4.80	$T_6$ 9.40
$T_6$ 5.20	$T_3$ 4.85	$T_4$ 6.60	$T_2$ 4.60	$T_5$ 7.00	$T_1$ 5.00
$T_4$ 4.25	$T_5$ 6.65	$T_6$ 9.30	$T_1$ 4.95	$T_3$ 9.30	$T_2$ 8.40

In the sequence, the data are analyzed (a) to identify the best treatment, (b) to test whether or not the average effect of  $T_3$  (100kg N/ha applied as urea) and  $T_4$  (100 kg N/ha) is same as the average effect of  $T_5$  ( $T_2$  + six insecticidal sprays) and  $T_6$  ( $T_4$  + six insecticidal sprays).

### 2.4.3 Analysis of data

We compute the following totals in Table 2.23.

**Table 2.23: Treatment, row and column totals**

Treatments Totals ( $T_i$ )	Rows Totals ( $R_j$ )	Columns Totals ( $C_k$ )
$T_1 = 1.75 + \dots + 4.95 = 22.20$	$R_1 = 3.10 + \dots + 5.30 = 26.35$	$C_1 = 3.10 + \dots + 4.25 = 26.75$
$T_2 = 3.85 + \dots + 8.40 = 28.40$	$R_2 = 4.80 + \dots + 5.40 = 25.85$	$C_2 = 5.95 + \dots + 6.65 = 28.90$
$T_3 = 3.10 + \dots + 9.30 = 34.20$	$R_3 = 3.00 + \dots + 7.10 = 33.45$	$C_3 = 1.75 + \dots + 9.30 = 31.45$
$T_4 = 5.30 + \dots + 4.25 = 31.60$	$R_4 = 6.40 + \dots + 9.40 = 36.75$	$C_4 = 6.40 + \dots + 4.95 = 34.40$
$T_5 = 6.40 + \dots + 6.65 = 38.55$	$R_5 = 5.20 + \dots + 5.00 = 33.25$	$C_5 = 3.85 + \dots + 9.30 = 36.40$
$T_6 = 5.95 + \dots + 9.30 = 43.55$	$R_6 = 4.25 + \dots + 8.40 = 42.85$	$C_6 = 5.30 + \dots + 8.40 = 40.60$

$$\text{Grand Total, } G = \sum_{i=1}^v T_i = \sum_{j=1}^v R_j = \sum_{k=1}^v C_k = 198.50$$

$$\text{Correction factor (CF)} = \frac{G^2}{v^2} = \frac{(198.5)^2}{(6)^2} = \frac{39402.25}{36} = 1094.506$$

$$\text{Treatments SS} = \frac{1}{v} \sum_{i=1}^v T_i^2 - CF$$

$$= \{(22.20)^2 + \dots + (43.55)^2\} / 6 - 1094.506 = 47.211$$

$$\begin{aligned} \text{Rows SS} &= \frac{1}{v} \sum_{j=1}^v R_j^2 - CF \\ &= \{(22.20)^2 + \dots + (43.55)^2\} / 6 - 1094.506 = 47.211 \end{aligned}$$

$$\begin{aligned} \text{Columns SS} &= \frac{1}{v} \sum_{k=1}^v C_k^2 - CF \\ &= \{(26.75)^2 + \dots + (40.6)^2\} / 6 - 1094.506 = 21.586 \end{aligned}$$

$$\begin{aligned} \text{Total SS} &= \sum_{i=1}^v \sum_{j=1}^v \sum_{k=1}^v y_{ijk}^2 - CF \\ &= (3.10)^2 + (5.95)^2 + \dots + (9.30)^2 + (8.40)^2 - 1094.506 = 128.333 \end{aligned}$$

$$\begin{aligned} \text{Error SS} &= \text{Total SS} - \text{Treatments SS} - \text{Rows SS} - \text{Columns SS} \\ &= 128.333 - 47.210 - 34.442 - 21.585 = 25.094 \end{aligned}$$

We now form the Analysis of Variance Table 2.24.

**Table 2.24: ANOVA table for cotton yield data**

Source	DF	SS	MS	F-value	Prob > F
Treatments	5	47.211	9.442	7.53	0.0004
Rows	5	34.442	6.883	5.49	0.0024
Columns	5	21.586	4.317	3.44	0.0210
Error	20	25.095	1.255		
Corrected Total	35	128.333			

From Table 2.24, one can easily see that the treatment effects are highly significant ( $p$ -value = 0.0004) meaning thereby that the null hypothesis of equal treatment effects is rejected. The treatments, therefore, influence the cotton yield. The rows and columns effects are also highly significant with respective  $p$ -values as 0.0024 and 0.0210. This is an evidence to the fact that the formation of rows and columns have been effective.

#### 2.4.4 Analysis using SAS

The design is a LSD with  $v = 6$  treatments and  $n = 36$  observations. In this design the number of rows is same as the number of columns, which in turn is same as the number of treatments. So in this design the replication number of treatments is equal to the number of treatments. The data has been analyzed using SAS. The commands and the data preparation are given in the sequel.

DATA lsd;

INPUT row col trt cyield;  
/\*the first column 'row' denotes the row number; the second column 'col' denotes the column number; the third column "trt" represents the treatment number and the last column 'cyield' represents the cotton yield\*/

CARDS;

1 1 3 3.10  
1 2 6 5.95  
1 3 1 1.75  
1 4 5 6.40  
1 5 2 3.85  
1 6 4 5.30  
2 1 2 4.80  
2 2 1 2.70  
2 3 3 3.30  
2 4 6 5.95  
2 5 4 3.70  
2 6 5 5.40  
3 1 1 3.00  
3 2 2 2.95  
3 3 5 6.70  
3 4 4 5.95  
3 5 6 7.75  
3 6 3 7.10  
4 1 5 6.40  
4 2 4 5.80  
4 3 2 3.80  
4 4 3 6.55  
4 5 1 4.80  
4 6 6 9.40  
5 1 6 5.20  
5 2 3 4.85  
5 3 4 6.60  
5 4 2 4.60  
5 5 5 7.00  
5 6 1 5.00  
6 1 4 4.25  
6 2 5 6.65  
6 3 6 9.30  
6 4 1 4.95  
6 5 3 9.30  
6 6 2 8.40  
;

```

PROC GLM data = lsd;
CLASS row col trt;
MODEL cyield = trt row col;
MEANS trt/tukey;
CONTRAST 'T3 T4 vs T5 T6' trt 0 0 1 1 -1 -1;
RUN;
    
```

### 2.4.5 Output of analysis

The results obtained from the analysis of data are described in the sequel.

**Table 2.25: Output using SAS**

#### ANOVA

Source	DF	SS	MS	F Value	Prob > F
Model	15	103.238	6.882	5.49	0.0003
Error	20	25.095	1.255		
Corrected Total	35	128.333			

R-Square	CV	Root MSE	cyield Mean
0.804	20.315	1.120	5.514

#### ANOVA

Source	DF	Type III SS	MS	F Value	Prob > F
Treatment	5	47.211	9.442	7.53	0.0004
Row	5	34.442	6.888	5.49	0.0024
Column	5	21.586	4.317	3.44	0.0210
Error	20	25.095	1.255		
Corrected Total	35	128.33			

It is worthwhile noting that the model with treatments effects, row effects and column effects explains about 80 per cent of the total variability in the data. As mentioned earlier also, this analysis of variance table divulges that the treatment effects are highly significant ( $p$ -value = 0.0004) meaning thereby that the null hypothesis of equal treatment effects is rejected. The rows and columns effects are also highly significant with respective  $p$ -values as 0.0024 and 0.0210. So running this experiment as a row-column design is justified and it is very apparent that there were two sources of variability in the experimental units.

The distribution of observations for each treatment is given Figure in 2.4. It is easily seen from the Figure also that the distribution of observations is very different for different treatments. The Figure clearly reveals that treatment number 3 is most variable and treatment number 5 is least variable.

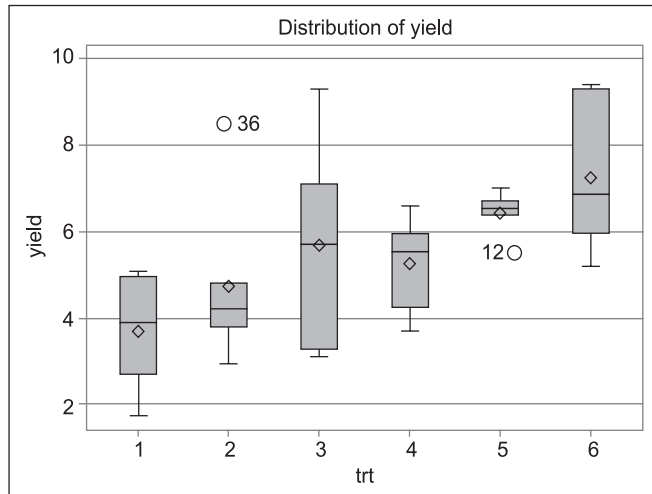


Figure 2.4: Treatment wise Box plot of seed yield

A pairwise comparison of treatment effects is made and the results are summarized in Table 2.26. Treatment 6 is the maximum yielding and is significantly different from treatment 1, which is the lowest yielding. Treatment 6 is, however, statistically at par with treatments 5, 3 and 4. Similarly, treatment 1 is statistically at par with treatments 2, 4 and 3. On the basis of yield, treatment 6 may be recommended as the best for cotton yield.

**Table 2.26: Treatments in descending order with letter display**

Means with the same letter are not significantly different					
Tukey Grouping		Mean	N	Treatment	
	A	7.258	6	6	
B	A	6.425	6	5	
B	A	C	5.700	6	3
B	A	C	5.267	6	4
B		C	4.733	6	2
		C	3.700	6	1

A comparison of treatments 3 and 4 with treatments 5 and 6 (null hypothesis that the average effect of treatments 3 and 4 is the same as the average effect of treatments 5 and 6) reveals that the difference is significant ( $p$ -value = 0.0076).

**Table 2.27: Contrast analysis result**

Contrast	DF	Contrast SS	Mean Square	F Value	Prob > F
3 4 vs 5 6	1	11.070	11.070	8.82	0.0076

The pairwise treatment comparisons can also be made without writing the treatments in descending order of the treatments LS Mean values. The results are given in Table 2.28.

**Table 2.28: Treatments with letter display**

Treatment Name	Treatment Description	Mean of 'cyield'	Rank of Treatment
1	Control	3.70 <sup>C</sup>	6
2	100kg N/ha applied as urea (half at final thinning and half at flowering as top dressing)	4.73 <sup>B,C</sup>	5
3	100kg N/ha applied as urea (80 kg N/ha in 4 equal split doses as spray and 20 kg N/ha at final thinning)	5.70 <sup>A,B,C</sup>	3
4	100 kg. N/ha applied as CAN (half at final thinning and half at flowering as top dressing)	5.27 <sup>A,B,C</sup>	4
5	$T_5 : T_2$ + six insecticidal sprays	6.43 <sup>A,B</sup>	2
6	$T_4$ + six insecticidal sprays	7.26 <sup>A</sup>	1
General Mean		5.51	
$SE(\hat{d})$		0.647	
Tukey HSD at 5%		2.033	

In Table 2.28, any two treatments whose Means have at least one letter common are not statistically significant using Fisher's Least Square Difference. Therefore, it follows that treatment 1 is significantly different from treatments 3, 4, 5 and 6, but is not significantly different from treatment 2. Similarly, treatment 6 is significantly different from treatments 1, 2, 3 and 4, but is not significantly different from treatment 5. Also treatment 2 is significantly different from treatments 5 and 6. Following the Table 2.28, it is also evident that treatment 3 is significantly different from treatments 1 and 6. Treatment 4, however, is not significantly different from treatments 2, 3 and 5.

Further, 100 kg. N/ha applied as CAN (half at final thinning and half at flowering as top dressing) along with six insecticidal sprays produces the maximum yield, though it is at par with 100kg N/ha applied as urea (half at final thinning and half at flowering as top dressing) coupled with six insecticidal sprays. It may be worthwhile mentioning here that all comparisons are made at 5 per cent level of significance.

### 2.4.6 Analysis using R

#### R code

```
d3=read.table("lsd.txt",header=TRUE)
attach(d3)
names(d3)
#Treatment means and standard deviations
aggregate(cyield, by=list(trt), mean)
aggregate(cyield, by=list(trt), sd)
#Treatment wise box plot of yield
boxplot(cyield~trt)
```



```

#set the contrast coefficients for testing average of treatments 3 and 4 with average of treatments
5 and 6
trt=factor(trt)
row=factor(row)
col=factor(col)
aov.out=aov(cyield~trt+row+col)
summary(aov.out)
library(lsmeans)
lsm <- lsmeans(aov.out, "trt")
contrast(lsm, list(con1 = c(0,0,1,1,-1,-1)))
#Tukey's honest significant difference test
TukeyHSD(aov.out, "trt")
#LSD
library(agricolae)
lsd.result <- LSD.test(aov.out, "trt")
lsd.result
detach(d3)

```

## 2.5 Conclusion

This Chapter has been devoted to introducing the basic designs like CRD, RCB design and Latin square design. SAS has been used for the analysis of data. The PROC GLM has been the major procedure used for analysis of data. The R code for the analysis of data has also been given.

It has been observed that in many experiments conducted as an RCB design (very few experiments are conducted as Latin square design), the block mean square is not high as compared to mean square error. In other words, block mean square is smaller than the error mean square. This is not a healthy situation. The basic purpose of forming blocks (or two systems of blocks as in Latin square design) is that there was variability in the experimental units. It is expected that the between blocks variability would be large and the within block variability would be small. But if the block effects are not significant, it means that substantial part of variability arising in the experimental units has not been accounted for by forming blocks. Obviously then the CV would also be large.

It may be re-emphasized that the variability in the experimental units is a very disturbing factor and it needs to be taken care of properly so as to enable a proper conduct of experiment.

For the benefit of the experimenters, a utility has been created at the "Design Resources Server" hosted at [www.iasri.res.in/design](http://www.iasri.res.in/design) to generate a randomized layout of these basic designs. There is also a provision for generating a data entry sheet based on the randomized plan either in TXT (Text file) or CSV (Comma Separated Values) formats. CSV/TXT files can be opened using any text editor or in MS<sup>®</sup>-Excel<sup>®</sup>. The experimenter may use "Datashet" hyperlink for downloading / opening generated datashet. Besides randomized layout, an outline of ANOVA is also shown for the benefit of the experimenters. The users may visit [http://iasri.res.in/design/Basic Designs/basicdesign.aspx](http://iasri.res.in/design/Basic%20Designs/basicdesign.aspx) and take advantage of this utility.