# DJS3B - DESIGN OF EXPERIMENTS

**Unit - I**

Fundamental principles of experiments – randomization, replication and local control. Size of experimental units. Analysis of variance- one-way and two-way classifications.

**Unit - II**

Analysis of Variance and Basic Designs: Concept of Cochran's Theorem. Completely randomized design(CRD)- Randomized block design(RBD) - Latin square design(LSD) and their analysis - Missing plot techniques in RBD and LSD.

**Unit - III**

Post ANOVA Tests: Multiple range test; Newman-Keul's test-Duncan's multiple range test-Tukey's test. Analysis of Covariance technique for RBD with one concomitant variable.

**Unit - IV**

Factorial experiments: $2^2$, $2^3$ and $2^n$ factorial experiments. Definitions and their analyses.

**Unit - V**

Principles of confounding –partial and complete confounding in $2^3$– balanced incomplete block design(BIBD)– parametric relationship of BIBD.

**REFERENCE BOOKS:**


1. Das, M.N. and Giri,N.C. (1988) Design and Analysis of Experiments($2^{nd}$ Edition). New Age International, New Delhi.
2. Douglas,C. Montgomery(2012) Design and Analysis of Experiemnts. John Wiley & sons, New York.
3. Goon A. M., Gupta, S.C. and Dasgupta, (2002)B. Fundamentals of Statistics, Vol.II, World Press, Kolkata.
4. Gupta, S. C. and V. K. Kapoor (1999) Fundamentals of Applied Statistics (Third Edition), Sultan Chand & Sons, New Delhi.
5. Dean, A and Voss (2006) Design and Analysis of Experiments. Springer India Private Limited, New Delhi.

Unit -I

## DESIGN 0F EXPERIMENTS

### 1.1 Introduction

In 1935 sir Ronald A. Fisher laid the foundation for the subject which has come to be known by the title of his book ' The Design of Experiments'. Since then the theory of experimental design has been considerably developed and extended. Applications of this theory are found today laboratories and research in natural sciences, engineering, and nearly all branches of social science.

### Definition

Design of experiments may be defined as the logical construction of the experiments in which the degree of uncertainty with which the inference is drawn may be well defined.

The subject matter of the design of experiments may includes;

1) Planning of the experiment.
2) Obtaining relevant information from it regarding statistical hypothesis under study, and
3) Making a statistical analysis of the data.

Allen L. Edwards the experimental design is called a randomized group design.

The experimenter may easily recognize three important phases of every project;

1) Experimental or planning phase.
   - i) Statement of problem.
   - ii) Choice of response or dependent variable.
   - iii) Selection of factors to be varied.
   - iv) Choice of levels of these factors.
     - Qualitative or quantitative.
       - Fixed or random.
     - How factor levels are to be combined.
2) Design phase,
   - i) Number of observations to be taken.
   - ii) Order of experimentation.
   - iii) Method of randomization to be used.
   - iv) Mathematical model to describe the experiment.

v) Hypothesis to be tested.

3) Analysis phase,

i) Data collection and processing.

ii) Computation of test statistics.

iii) Interpretation of results for the experiment.

## 1.2 Definitions:

**1. Experiment**

An experiment is a device or a means of getting an answer to the problem under consideration.

Experiment can be classified into two categories;

i) Absolute

ii) Comparative

**i) *Absolute experiment***

Absolute experiments consist in determining the absolute value of some characteristics like,

a) Obtaining average intelligence quotient (I.Q) of a group of people.

b) Finding the correlation co-efficient between two variables in a bivariate distribution etc.

**ii) Comparative experiment**

Comparative experiments are designed to

Compare the effect of two or more objects on some population characteristics.

Example;

- Comparison of different fertilizers.
- Different kinds of verities of a crop.
- Different cultivation processes etc.,

**2. Treatments**

Various objects of comparison in a comparative experiment are termed as treatments.

Example

In field experimentation different fertilizers or different varies of crop or different methods cultivation are the treatments.

**3. Experimental unit**

The smallest division of the experimental material to which we apply the treatments and on which we make observations on the variable under study.

Example

i) In field experiments the plot of land is the experimental unit. In other experiments, unit may be a patient in a hospital, a lump of dough or a batch of seeds.

**4. Blocks**

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

**5. Yield**

The measurement of the variable under study on different experimental units are termed as yields.

**6. Experimental error**

Let us suppose that a large homogeneous field is divided into different plots (of equal shape and size) and different treatments are applied to these plots. If the yields from some of the treatments are more than those of others, the experimenter is faced with the problem of deciding if the observed differences are really due to treatment effects or they are due to chance (uncontrolled) factors. In field experimentation, it is a common experience that the fertility gradient of the soil does not follow any systematic pattern but behaves in an erratic fashion. Experience tells us that even if the same is used in all the plots, the yields would still vary due to the differences in soil fertility. Such variation from plot to plot, which is due to random factors beyond human control, is spoken of as experimental error.

**7. Replication**

Replication means the execution of an treatments more than once. In other words, the repetition of treatments under investigation is known as replication.

**8. Precision**

The reciprocal of the variance of the mean is termed as the precision. Thus for an experiment replicated r times is given by.

$$\frac{1}{\text{var}(\bar{x})} = r/\sigma^2$$

Where $\sigma^2$ is the error variance per unit.

**9. Efficiency of a Design**

Consider the designs $D_1$ and $D_2$ with error variances per unit $\sigma_1^2$ and $\sigma_2^2$ and replications $r_1$ and $r_2$ respectively. Then the variance of the difference between two treatment means is given by

$2\sigma_1^2/r_1$ *and* $2\sigma_2^2/r_2$ for $D_1$ and $D_2$ respectively. Then the ratio

$E = \frac{2\sigma_2^2}{r_2} \times \frac{r_1}{2\sigma_1^2} = \frac{r_1}{\sigma_1^2} \div \frac{r_2}{\sigma_2^2}$ is termed as efficiency of design $D_1$ w.r.t $D_2$.

**10. Uniformity Trials**

The fertility of the soil does not increase or decrease uniformity in any direction but is distributed over the entire field in an erratic manner. Uniformity trails enable us to have an idea about the fertility variation of the field. By uniformity trail, we mean a trail in which the field (experimental material) is divided into small units (plots) and the same treatment is applied on each of the units and their yields are recorded.

## 1.3 Basic Principles of Experimental Designs

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

The principles of experimental design;-

1) Replication
2) Randomization
3) Local control

### 1) Replication

Replication means the repetition of the treatments under investigation. An experimenter resorts to replication in order to average out the influence of the chance factors on different experimental units. Thus, the repetition of treatment results is more reliable estimate than is possible with a single observation

Advantages of replication

1. Replication serves to reduce experimental error and thus enables us to obtain more precise estimates of the treatment effects.
2. From statistical theory we know that the standard Error (S.E) of the mean of a sample size n is $\sigma/\sqrt{n}$, where σ is the standard deviation of the population. Thus if a treatment is replicated r times, then the S.E of its mean effect is $\sigma/\sqrt{n}$, where $\sigma^2$ is the variance of the individual plot is estimated from error variance. Thus " the precision of the experiment is inversely proportional to the square of the Replication has an important but limited role in increasing the efficiency of the design.

### 2) Randomization

We have seen that replication will provide an estimate of experimental error. For valid conclusions about our experimental results, we should have not merely an estimate of experimental error but it should be an unbiased estimate. Also, if our conclusions are to be valid, the treatment means and also differences among treatment means should be estimated without any bias. For the purpose we use the technique of randomization.

When all the treatments have equal chances of being allocated to different experimental units it is known as randomization.

The following are the main objectives of randomization.

i) The validity of the statistical test of the Significance.

i.e.) t-test for testing the significance of the difference of two means. F-test for testing the homogeneity of variance.

ii) The purpose of randomness is to assure that the source of variation, not controlled in the experiment operate randomly. Randomization eliminates bias in any form.

**3) Local control**

We know that the estimate of experimental error is based on the variations from experimental unit to experimental unit. In other words, the error in an experiment is a measure of " within block" variation. This suggests that if we group the homogeneous experimental units into blocks, the experimental error will be reduced considerably. If the experimental material, say field for agriculture experimentation is heterogeneous and different treatment are allocated to various units at random over the entire field the soil heterogeneous will also enter the uncontrolled factors and thus increase the experimented error. It is desirable to reduce the experimental error as for as practicable without unduly increasing the number of replications, so that even smaller difference between treatments can be detected as significant.

The process of reducing the experimental error by dividing relatively heterogeneous experimental area (field) into homogeneous blocks is known as local control.

**Remarks:**

1. Local control, by reducing the experimental error, increases the efficiency of the design.
2. Various forms of arranging the units(plots) into homogeneous groups(blocks) have so far been evolved and are known as experimental designs, e.g., Randomised Block Design, Latin Square Design etc.,

**1.4 Analysis of Variance**

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

inherent in nature. The total variation in any set of numerical data is due to number of causes which may be classified as: (i) Assignable causes, and (ii) Chance causes.

The variation due to assignable causes can be detected and measured whereas the variation due to chance causes is beyond the control of human hand cannot be traced separately.

**Definition.** According to Prof. R.A. Fisher, Analysis of variance (ANOVA) is the "Separation of variance ascribable to one group of causes from the variance ascribable to other group."

**Assumptions for ANOVA Test.**

ANOVA test is based on the test statistics F (or Variance Ratio).

For the validity of the F-test in ANOVA, the following assumptions are made:

(i) The observations are independent,

(ii) Parent population from which observations are taken is normal, and

(iii) Various treatment and environmental effects are additive in nature.

In the following sequence we will discuss the analysis of variation for:

(a) One-way classification, and (b) Two-way classification.

## 1.5 ONE-WAY CLASSIFICATION

Let us suppose that N observations $y_{ij}$, (i=1, 2... k; j= 1, 2, .....,$n_i$) of a random variable Y are grouped, on some basis, into k classes of sizes $n_1$, $n_2$, ...., $n_k$ respectively, ( $N = \sum_{i=1}^{k} n_i$ ) as exhibited in table

Table 1.1: **ONE-WAY CLASSFIED DATA**

| Class | Sample Observations | Total | Mean |
|---|---|---|---|
| 1 | $y_{11}y_{12} \quad \cdots \quad y_{1n_1}$ | $T_{1.}$ | $\overline{y_{1.}}$ |
| 2 | $y_{21}y_{22} \quad \cdots \quad y_{2n_1}$ | $T_{2.}$ | $\overline{y_{2.}}$ |
| . | .... | . | . |
| . | .... | . | . |
| . | .... | . | . |
| i | $y_{i1}y_{i2} \quad \cdots \quad y_{in_i}$ | $T_{i.}$ | $\overline{y_{i.}}$ |
| . | .... | . | . |
| . | .... | . | . |
| . | .... | . | . |
| k | $y_{k1}y_{k2} \quad \cdots \quad y_{kn_k}$ | $T_{k.}$ | $\overline{y_{k.}}$ |

The total variation in the observation $y_{ij}$ can be split into the following two components:

(i) The variation between the classes or the variation due to different bases of classification, commonly known as treatments.

(ii) The variation within the classes, i.e, the inherent variation of the random variable within the observations of a class.

The firs type of variation is due to assignable causes which are beyond the control of human hand.

The main objective of analysis of variance technique is to examine if there is significant difference between the classes means in view of the inherent variability within the separate classes.

In particular, let us consider the effect of k different rations on the yield in milk of N cows (of the same breed and stock) divided into k classes of sizes

$n_1, n_2, \ldots, n_k$ respectively, $N = \sum_{i=1}^{k} n_i$ . Here the sources of variation are:(i) Effect of the ration (treatment) : $t_i$; i= 1, 2, ....,k.

(ii) Error (ε) produced by numerous causes of such magnitude that they are not detected and identified with the knowledge that we have and they together produce a variation of random nature obeying Gaussian (Normal) law of errors.

### 1.5.1 Analysis of one way Classified Data

Let $y_{ij}$ denote the $j^{th}$ observations in the $i^{th}$ level of a factor A and let $y_{ij}$ be corresponding random variable. Let the mathematical model for one way classified data

$$y_{ij} = \mu + t_i + e_{ij}; \quad i = 1,2,\ldots,k; \quad j = 1,2,\ldots,n_i$$

Where μ is the general mean effect

$T_i$ is the effect $i^{th}$ level of factor A

$$e_{ij} \overset{i.i.d}{\sim} N(0,\sigma_e^2)$$

$E(y_{ij}) = \mu + t_i$

μ and $t_i$ , i=1,2,…,k can be estimated by least square method that is minimizing error sum of squares

$$E(e'e)^2 = \sum_{ij}\left(y_{ij} - E(y_{ij})\right)^2$$

$$= \sum_{ij}\left(y_{ij} - (\mu + t_i)\right)^2$$

$$= \sum_{ij}\left(y_{ij} - \mu - t_i\right)^2$$

$$\frac{\partial E(e'e)}{\partial \mu} = 0$$

$$2\sum_{ij}\left(y_{ij} - \mu - t_i\right)(-1) = 0$$

$$\sum_{ij}\left(y_{ij} - \mu - t_i\right) = 0$$

$$= \sum_{ij} y_{ij} - \sum_{ij}\mu - \sum_{ij} t_i = 0$$

$$\sum_{ij} y_{ij} = \sum_{ij} \mu + \sum_{ij} t_i$$

$$\sum_{ij} y_{ij} = G \text{ =Grand Total}$$

$$G = \sum_{i} n_i \mu + \sum_{i} n_i t_i$$

$$\sum_{i} n_i = n$$

$$G = n\mu + \sum_{i} n_i t_i \quad \ldots(1)$$

$$\frac{\partial (e'e)}{\partial t_i} = 0$$

$$2\sum_{j} \left(y_{ij} - \mu - t_i\right)(-1) = 0$$

$$\sum_{j} \left(y_{ij} - \mu - t_i\right) = 0$$

$$= \sum_{j} y_{ij} - \sum_{j} \mu - \sum_{j} t_i = 0$$

$$\sum_{j} y_{ij} = \sum_{j} \mu + \sum_{j} t_i$$

$$\sum_{j} y_{ij} = T_i$$

$$T_i = n_i \mu + n_i t_i \quad \ldots(2)$$

Equation (1) and (2) are not independent

We assume that

$$\sum_{i} n_i t_i = 0$$

From equation (1)

G=nμ

$$\hat{\mu} = \frac{G}{n}$$

From equation (2)

$$T_i = n_i\mu + n_i t_i$$

$$T_i = n_i \frac{G}{n} + n_i t_i$$

$$T_i - n_i \frac{G}{n} = n_i t_i$$

$$\hat{t}_i = \frac{T_i}{n_i} - \frac{G}{n}$$

Error Sum of Squares

$$E(e'e)^2 = \sum_{ij} (y_{ij} - \hat{\mu} - \hat{t}_i)^2$$

$$= \sum_{ij} (y_{ij} - \mu - t_i)(y_{ij} - \mu - t_i)$$

$$= \sum_{ij} y_{ij} (y_{ij} - \mu - t_i) + other\ terms\ are\ vanished$$

$$= \sum_{ij} \left[ y_{ij}^2 - \hat{\mu} y_{ij} - \hat{t}_i y_{ij} \right]$$

$$= \sum_{ij} y_{ij}^2 - \hat{\mu} G - \sum_i t_i \sum_j y_{ij}$$

$$= \sum_{ij} y_{ij}^2 - \frac{G^2}{n} - \left( \sum_i \frac{T_i^2}{n_i} - \frac{G^2}{n} \right)$$

$$= \left( \sum_{ij} y_{ij}^2 - \frac{G^2}{n} \right) - \left( \sum_i \frac{T_i^2}{n_i} - \frac{G^2}{n} \right)$$

Error Sum of Square (E.S.S) = Total Sum of Square (T.S.S)-Treatment Sum of Square (Tr.S.S)

**Table 1.2: Anova Table for One –way Classified Data**

| Source of variation | d.f | Sum of squares | Mean sum of squares | F-ratio |
|---|---|---|---|---|
| Treatment(Ration) | k-1 | $S_t^2$ | $s_t^2 = \frac{S_t^2}{(k-1)}$ | $F = \frac{s_t^2}{s_E^2} = F_{k-1,n-k}$ |
| Error | n-k | $S_E^2$ | $s_E^2 = \frac{S_t^2}{(n-k)}$ | |
| Total | n-1 | $S_T^2$ | | |

Under the null hypothesis, $H_0 = t_1 = t_2 = \cdots = t_k$ against the alternative that all t's are not equal, the test statistic $F = \frac{s_t^2}{s_E^2} = F_{k-1,n-k}$

i.e., F follows F (central) distribution with (k-1, n-k) d.f

If $F > F_{(k-1,\ n-k)}$ (α) then $H_0$ is rejected at α % level of significance and we conclude that treatments differ significantly. Otherwise $H_0$ accepted.

**Problem 1.1.**

The average number of days survived by mice inoculated with 5 strains of typhoid organisms along with their standard deviation and number of mice involved in each experiment is given below. On the basis of these data, what would be your conclusions regarding the strains of typhoid organisms?

| Strains of typhoid | A | B | C | D | E |
|---|---|---|---|---|---|
| No. of mice, $n_i$ | 10 | 6 | 8 | 11 | 5 |
| Average, $\overline{y_i}$ | 10.9 | 13.5 | 11.5 | 11.2 | 15.4 |
| Standard deviation, $s_i$ | 12.72 | 5.96 | 3.24 | 5.65 | 3.64 |

Solution.

Here we set up the Null Hypothesis, $H_o$: Different strains of typhoid organisms are homogeneous,

i.e.,

$H_o$:$\mu_A = \mu_B = \mu_C = \mu_D =$

$\mu_E H_1$: *At least two of the means are different*

Let $T_{i.}$ Be total for the i[th] strain of typhoid and $G = \sum_i T_i$ be the grand total.

Then $\overline{y}_{i.} = \frac{T_{i.}}{n_i} \Rightarrow T_{i.} = n_i \overline{y}_{i.}$

Also $\Rightarrow s_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}^2 - \overline{y}_i^2 \Rightarrow \sum_{j=1}^{n_i} y_{ij}^2 = n_i\left(s_i^2 + \overline{y}_i^2\right)$

Which gives the S.S of observations for the i[th] strain typhoid.

CALCULATIONS FOR VARIOUS SUM OF SQUARES

$$C.F = \frac{G^2}{N} = \frac{(482.2)^2}{40} = 5{,}812.92$$

$$R.S.S = \sum_i \sum_j y_{ij}^2 = \sum_i n_i\left(s_i^2 + \overline{y}_i^2\right) == 8237.73$$

$$T.S.S = R.S.S - C.F. = 8{,}237.73 - 5{,}812.92 = 2{,}424.81$$

S.S due to strains of typhoid

$$= \sum_i \frac{T_i^2}{n_i} - C.F.$$

$$= \frac{(109)^2}{10} + \frac{(81)^2}{6} + \frac{(92)^2}{8} + \frac{(123.2)^2}{11} + \frac{(77)^2}{5} - 5812.92$$

$$= 1{,}188.1 + 1{,}093.5 + 1{,}058 + 1{,}379.84 + 1{,}185.8 - 5{,}812.49$$

$$Error\ S.S. = T.S.S. - S.S.\ due\ to\ atrain = 2{,}424.81 - 92.32 = 2{,}332.49$$

Table 1.3 ANOVA TABLE

| Sources of Variatoin | d.f | Sum of squares | Mean S.S | V.R.(F) |
|---|---|---|---|---|
| Between strains of typhoid | 4 | 92.32 | $\frac{92.32}{4} = 23.08$ | $\frac{66.63}{23.08} = 2.89$ |
| Error | 35 | 2,332.49 | $\frac{2{,}332.49}{35} = 66.63$ | |
| Total | 39 | 2,424.81 | | |

Tabulated $F_{0.05}$ and 35 d.f = 5.735. Since calculated value of F is less than the tabulated value, it is not significant at 5% level of significance and the null hypothesis $H_o$ may be accepted.

## 1.6 TWO-WAY CLASSIFICATION (ONE OBSERVATION PER CELL)

Suppose n observations are classified into k categories (or classes), say $A_1$, $A_2$, ....,$A_k$ according to some criterion, A: and into h categories, say,

$B_1$, $B_2$, ...$B_h$ according to some criterion B, having kh combinations

($A_i$,$B_j$) i=1, 2, ...., k ; j= 1, 2, ..., h; often called cells. This scheme of classification according to two factors or criteria is called two-way classification and its analysis is called two-way analysis of variance. The number of observations in each cell may be equal or different, but we shall consider the case of one observation per cell so that n=hk, i.e., the total number of cells is n=nk.

In the two-way classification, the values of the response variable are affected by two factors.

For example, the yield of milk may be affected by differences in treatments, (i.e, rations as well as the differences in variety, i.e., breed and stock of the cows. Let us suppose that the n cows are divided into h different groups or classes according to their breed and rations given at random to cows in each group) on the yield of milk.

Let $y_{ij}$ = [Yield of milk from the cow of $j^{th}$ breed or stock, fed on the ration i]; i= 1, 2, ..., k; j= 1, 2, ..., h

Note that the suffix i refers to the treatments (rations) and the suffix j refers to the breed and stock of the cow. The yield can be expressed as variable values in the following k × h two-way. One factor of variation, say, varieties (breed and stock of cows) is represented along the columns and the other factor of variation, say, treatments (rations) is represented along the rows of the table.

**Table 1.4: TWO-WAY CLASSIFIED DATA**

| Treatments (Rations) | Varieties of Cows<br>1 2 … j … h | Row Totals<br>= ( $\sum_j y_{ij}$ ) | Row Means<br>= ( $\sum_j y_{uj}$ )/h |
|---|---|---|---|
| 1 | $y_{11}y_{12}$ … $y_{1j}$ … $y_{1h}$ | $T_{1.}$ | $\bar{y}_{1.}$ |
| 2 | $y_{21}y_{22}$ … $y_{2j}$ … $y_{2h}$ | $T_{2.}$ | $\bar{y}_{2.}$ |
| . | ..... | . | . |
| . | ..... | . | . |
| i | ..... | | |
| . | $y_{i1}y_{i2}$ … $y_{ij}$ … $y_{ih}$ | . | . |
| . | ..... | . | . |
| K | ..... | $T_{k.}$ | $\bar{y}_{k.}$ |
| | $y_{k1}y_{k2}$ … $y_{kj}$ … $y_{kh}$ | | |

| | | | |
|---|---|---|---|
| Column Totals | $T_{1.}T_{.2}$ … $T_{.j}$ … $T_{.h}$ | G= $\sum\sum y_{ij}$ | |
| Column means $=(\sum_i y_{ij})/k$ | $\bar{y}_{.1}\bar{y}_{.2}$ … $\bar{y}_{.j}$ … $\bar{y}_{.h}$ | | |

### 1.6.1. Analysis Two Way Classified Data

The appropriate model of this data is

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}; \qquad i = 1,2,\ldots,p; \quad j = 1,2,\ldots,q$$

Where $y_{ij}$ is the yield of the $(i,j)^{th}$ element which is in the $i^{th}$ row and $j^{th}$ column with p levels and q levels, where μ is the general mean effect

α= $\alpha_i$ ($i^{th}$ row effect)

β= $\beta_j$ ($j^{th}$ column effect)

$$e_{ij} \overset{i.i.d}{\sim} N(0,\sigma_e^2)$$

μ, $\alpha_i$ and $\beta_j$ are estimated by the method of least squares

$$E = \sum_{ij} e_{ij} = \sum_{ij} \left(y_{ij} - \mu - \alpha_i - \beta_j\right)$$

$$E = \sum_{ij} e_{ij}^2 = \sum_{ij} \left(y_{ij} - \mu - \alpha_i - \beta_j\right)^2 \ldots(1)$$

$$\frac{\partial E}{\partial \mu} = 0; \ \frac{\partial E}{\partial \alpha_i} = 0; \ \ \frac{\partial E}{\partial \beta_j} = 0$$

Differentiate w.r.to μ

$$2\sum_{ij} \left(y_{ij} - \mu - \alpha_i - \beta_j\right)(-1) = 0$$

$$\sum_{ij}\left(y_{ij}-\mu-\alpha_i-\beta_j\right)=0$$

$$\sum_{ij}y_{ij}-\sum_{ij}\mu-\sum_{j}\sum_{i}\alpha_i-\sum_{ij}\beta_j=0$$

where $\sum_{ij}y_{ij}=G$

$$G=pq\mu+q\sum_{i}\alpha_i+p\sum_{j}\beta_j \quad \ldots(2)$$

Differentiate w.r .to $\alpha_i$ in equation (1 )

$$=2\sum_{j}\left(y_{ij}-\mu-\alpha_i-\beta_j\right)(-1)=0$$

$$=\sum_{j}\left(y_{ij}-\mu-\alpha_i-\beta_j\right)=0$$

$$=\sum_{j}y_{ij}-\sum_{j}\mu-\sum_{j}\alpha_i-\sum_{j}\beta_j=0$$

where $\sum_{j}y_{ij}=T_i$

$$T_i=q\mu+q\alpha_i+\sum_{j}\beta_j \quad \ldots(3)$$

Differentiate w.r .to $\beta_j$ in equation (1 )

$$=2\sum_{i}\left(y_{ij}-\mu-\alpha_i-\beta_j\right)(-1)=0$$

$$=\sum_{i}\left(y_{ij}-\mu-\alpha_i-\beta_j\right)=0$$

$$=\sum_{i}y_{ij}-\sum_{i}\mu-\sum_{i}\alpha_i-\sum_{i}\beta_j=0$$

where $\sum_{i}y_{ij}=T_j$

$$T_j=p\mu+\sum_{i}\alpha_i+p\beta_j \quad \ldots(4)$$

We assume that $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$

From equation (2)

G=pqμ

$$\frac{G}{pq} = \hat{\mu}$$

From equation (3)

$T_i = q\mu + q\alpha_i$

$T_i - q\mu = q\alpha_i$

$$\frac{T_i - q\mu}{q} = \hat{\alpha}_i$$

$$\hat{\alpha}_i = \frac{T_i}{q} - \frac{G}{pq}$$

From equation (4)

$$T_j = p\mu + + p\beta_j$$

$$\hat{\beta}_j = \frac{T_j}{p} - \frac{G}{pq}$$

Then the error sum of squares,

$$E = \sum_{ij} (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

$$E = \sum_{ij} (y_{ij} - \mu - \alpha_i - \beta_j)(y_{ij} - \mu - \alpha_i - \beta_j)$$

$$E = \sum_{ij} y_{ij} (y_{ij} - \mu - \alpha_i - \beta_j) + other\ terms\ are\ vanished$$

$$= \sum_{ij} y_{ij}^2 - \hat{\mu} \sum_{ij} y_{ij} - \sum_{ij} \hat{\alpha}_i y_{ij} \sum_{ij} y_{ij} \hat{\beta}_j$$

$$= \sum_{ij} y_{ij}^2 - \frac{G}{pq} \sum_{ij} y_{ij} - \sum_{ij} y_{ij} \left( \frac{T_i}{q} - \frac{G}{pq} \right) - \sum_{ij} y_{ij} \left( \frac{T_j}{p} - \frac{G}{pq} \right)$$

Where $\sum_{ij} y_{ij} = G, \sum_{j} y_{ij} = T_i \quad , \sum_{i} y_{ij} = T_j$

$$= \left( \sum_{ij} y_{ij}^2 - \frac{G^2}{pq} \right) - \left( \frac{\sum_i T_i^2}{q} - \frac{G^2}{pq} \right) - \left( \frac{\sum_j T_j^2}{p} - \frac{G^2}{pq} \right)$$

= Total Sum of Square (T.S.S) = Row Sum of Square (R.S.S)-Column Sum of Square (C.S.S)

**Table 1.5: ANOVA TABLE FOR TWO-WAY DATA WITH ONE OBSERVATION PER CELL RANDOM EFFECT MODEL**

| Sources of variation | S.S | d.f | M.S.S | Variance ratio |
|---|---|---|---|---|
| Factor A | S.S.A. | p-1 | $M.S.A. = \frac{S.S.A.}{p-1}$ | |
| Factor B | S.S.B. | q-1 | $M.S.B. = \frac{S.S.B.}{q-1}$ | $F_A = \frac{M.S.A.}{M.S.E}$ |
| Error | S.S.E. | (p-1) (q-1) | $M.S.B. = \frac{S.S.B.}{(p-1)(q-1)}$ | $F_B = \frac{M.S.B.}{M.S.E}$ |
| Total | T.S.S. | pq-1 | | |

Under the null hypothesis $H_{0t} = t_1 = t_2 = \ldots = t_p$ against the alternative that all t's are not equal the test statistic is :

$$F_A = \frac{M.S.A}{M.S.E} \sim F_{[(p-1),(p-1)(q-1)]}$$

i.e., $F_T$ follows F(central) distribution with [(p-1), (p-1)(q-1)] d.f. Thus if $F_A$ is greater than tabulated F for [(p-1), (p-1)(q-1)] d.f, at certain level of significance, usually 5 % then we reject the null hypothesis$H_{0t}$ and conclude that the

treatments differ significantly. If $F_t$ is less than tabulated value then $F_A$ is not significant and we conclude that the data do not provide any evidence against the null hypothesis which may be accepted.

Similarly under the null hypothesis $H_{ob}=b_1=b_2=\ldots=b_q$, against the alternative that b' s are not equal, the test statistics is:

$$F_B = \frac{M.S.B}{M.S.E} \sim F_{[(q-1),(p-1)(q-1)]}$$

And we discuss its significance as explained above.

**Problem 1.2.**

Three different methods of analysis $M_1$, $M_2$, $M_3$ are used to determine of a certain constituent in the sample. Each method is used by five analysis in the results, and the results are given in the results are given in table

| | Method | | |
|---|---|---|---|
| Analyst | M1 | M2 | M3 |
| 1 | 7.5 | 7.0 | 7.1 |
| 2 | 7.4 | 7.2 | 6.7 |
| 3 | 7.3 | 7.0 | 6.9 |
| 4 | 7.6 | 7.2 | 6.8 |
| 5 | 7.4 | 7.1 | 6.9 |

Do these results indicate a significant variation either between the methods or between the analysts?

Solution:

Here two factors of variation are, say,

A : Analysts, represented along the rows of the Table.

B : Methods, represented along the columns of the table.

**Null hypothesis :**

$H_{0A}$ : $\mu_{1.} = \mu_{3.} = \mu_{4.} = \mu_{5.}$ , i.e., there is no significant difference between the analysts.

$H_{oB}$ : $\mu_{.1} = \mu_{,2} = \mu_{.3}$, i.e., there is no significant difference between the methods.

**Alternative Hypothesis :**

$H_{1A}$ : At least two of $\mu_{1.}, \mu_{2.}, \ldots, \mu_{5.}$ are different.

$H_{1B}$ : At least two of $\mu_{,1}, \mu_{.2}, \ldots, \mu_{5}$ are different.

$y_i$ = response of the $i^{th}$ analysts and the $j^{th}$ methods ($i = 1,2, \ldots ,5$ ; $j = 1,2,3$)

In the usual notations, we have

K=5, h=3 and N=h×k = 3×5 = 15

Table 1.6: CALCULATIONS FOR VARIOUS S.S

| Analyst | Method | | | $T_{i.} = \sum_j y_{ij}$ | $T_{i.}^2$ |
|---|---|---|---|---|---|
| | $M_1$ | $M_2$ | $M_3$ | | |
| | | | | 21.6 | 466.56 |
| 1 | 7.5 | 7.0 | 7.1 | 21.3 | 453.69 |
| 2 | 6.7 | 7.4 | 7.2 | 21.2 | 449.44 |
| 3 | 6.9 | 7.3 | 7.0 | 21.6 | 466.56 |
| 4 | 6.8 | 7.6 | 7.2 | 21.4 | 457.96 |

| | | | | | |
|---|---|---|---|---|---|
| 5 | 6.9 | 7.4 | 7.1 | | |
| Column total, $T_{.j}$ | 37.2 | 35.5 | 34.4 | G = 107.1 | $\sum_i T_i^2$ = 2,294.21 |
| $T_{.j}^2$ | 1,383.84 | 1,260.25 | 1,183.36 | $\sum_j T_{.j}^2$ = 3,827.45 | |

Raw S.S. (R.S.S.) = $\sum_i \sum_j y_{ij}^2 = (7.5)^2 + (7.0)^2 + \cdots + (7.1)^2 + (6.9)^2$

$= 155.66 + 151.49 + 149.90 + 155.84 + 152.78$ $(Raw - wise\ S.S.)$

$= 765.67$

Correction Factor = $\frac{G^2}{N} = \frac{(107.1)^2}{15} = \frac{11,470.41}{15}$

Total S.S. = R.S.S. – $C.F.$ = 765.67 – 764.694 = 0.976

S.S.A. = S.S. due to factor A (Analysis) = $\frac{1}{\mathrm{h}} \sum_i T_{i.}^2 - CF$

$= \frac{2,294.21}{3} - 764.694 = 764.737 - 764.694 = 0.043$

S.S.B. = S.S. due to factor B (methods) = $\frac{1}{\mathrm{k}} \sum_i T_{.j}^2 - CF$

$= \frac{3,827.45}{5} - 764.694 = 765.49 - 764.694 = 0.796$

S.S. due to Error (S.S.E.) = T.S.S. – S.S.A. – S.S.B. = 0.976 – 0.043 – 0.796 = 0.137

Table 1.7: ANOVA TABLE

| Sources of variation | d.f | S.S. | Mean SS (MSS) | Variance Ratio |
|---|---|---|---|---|
| (1) | (2) | (3) | (4)=(3)/(2) | (F) |
| Factor A (Analysts) | k - 1 = 5 – 1 = 4 | 0.043 | 0.0108 | $F_A = \frac{0.0108}{0.0171} < 1$ |
| Factor B (Methods) | h – 1 = 3 – 1 = 2 | 0.796 | 0.3980 | $F_A = \frac{0.3980}{0.0171} = 23.27^*$ |
| Error | 4×2=8<br>OR<br>14 - (4+2) = 8 | 0.137 | 0.0171 | |
| Total | N – 1 =15 – 1 =14 | 0.976 | | |

Tabulated $F_{0.05}$ (2,8) = 19.40.

Since the calculated value $F_A < 1$, it is not significant and we fail to reject $H_{OA}$. Hence, there is no significant difference between the analysis.

Since the calculated value of $F_B$ = 23.27 is greater than the tabulated value, it is significant. Hence the hypothesis $H_{OB}$ of the homogeneity of the methods is rejected at 5% level of significance thus, we conclude that the methods differ significantly at 5% level of significance.

**Unit –II**

## 2.1 Analysis of Variance

The analysis of variance is a powerful statistical tool tests of significance. The test of significance based on t-distribution is an adequate procedure only for testing the significance of the difference between two. In a situation when we have three or more samples to consider at a time an alternative procedure is needed for testing the hypothesis that all the samples are drawn from the same population, i.e., they have the same mean. For example, five fertilizers are applied to four plots each of wheat and yield of wheat on each of the plot is given. We may be interested in finding out whether the effect of these fertilizers on the yield is significantly different or in other words, whether the samples have come from the same normal population. The answer to this problem is provided by the technique of analysis of variance. The basic purpose of the analysis of variance is test the homogeneity of several means.

## 2.2 Cochran's Theorem

Let $X_1, X_2, \ldots, X_n$, denote a random sample from normal population $N(0, \sigma^2)$. Let the sum of the squares of these values be written in the form:

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \cdots . + Q_k$$

Where $Q_j$ is a quadratic from in $X_1, X_2, \ldots, X_n$, with rank (degrees of freedom) $r_j$: j= 1, 2, …., k. Then the random variables $Q_1, Q_2, \ldots, Q_k$ are mutually independent and $Q_j/\sigma^2$ is $\chi^2$-variate with $r_j$ degrees of freedom if and only if

$$\sum_{j=1} r_j = n .$$

## 2.3 Completely Randomised Design(CRD)

In this design the experimental units are allotted at random to the treatments, so that every unit gets the same chance of receiving every treatment.

For example

Let there be five treatments each to be replicated four times. There are, therefore, 20 plots. Let these plots be numbered from 1 to 20 conveniently.

When a coin is tossed, there are two events, that is, either the head comes up, or the tail. We denote the "head" by H and the "tail" by T.

**Layout of CRD**

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| A | C | A | D |
| 5 | 6 | 7 | 8 |
| B | D | B | D |
| 9 | 10 | 11 | 12 |
| C | B | C | A |
| 13 | 14 | 15 | 16 |
| B | D | A | C |

**Advantages of CRD**

i) It is easy to layout the design.
ii) It results in the maximum use of the experimental units since all the experimental materials can be used.
iii) It allows complete flexibility as any number of treatments and replicates may be used. The number of replicates , if desired, can be varied from treatment to treatment.
iv) The statistical analysis is easy even if the number of replicates are not the same for all treatments
v) It provides the maximum number of degrees of freedom for the estimation of the error variance, which increases the sensitivity or the precision of the experiment for small experiments.

**Disadvantages of CRD**

i) In certain circumstances, the design suffers from the disadvantage of being inherently less informative than other more sophisticated layouts. This usually happens if the experimental material is not homogeneous.
ii) Since randomisation is not restricted in any direction to ensure that the units receiving one treatment are similar to those of receiving other treatment, the whole variations among the experimental units is included in the residual variance.

iii) This makes the design less efficient and results in less sensitivity in detecting significant effects.

**Applications**

Completely randomized design is most useful in laboratory technique and methodological studies, e.g., in physics, chemistry, in chemical and biological experiments , in some green house studies, etc.,

### 2.3.1 Statistical Analysis of CRD

The model is

$$y_{ij} = \mu + t_i + e_{ij} \qquad i = 1,2,...,k; \qquad j = 1,2,..n_i$$

Where $y_{ij}$ is the yield

$\mu$ is the general mean effect

$t_i$ is the treatment effect

$e_{ij}$ is the error term mean zero and variance $\sigma^2$

$E(y_{ij}) = \mu + t_i$ , i= 1, 2 …,k can be estimated by method of least square that is minimizing error sum of square

$$E(e'e)^2 = \sum_{ij}\left(y_{ij} - E(y_{ij})^2\right)$$

$$= \sum_{ij}\left(y_{ij} - (\mu + t_i)\right)^2$$

$$\frac{\partial(e'e)}{\partial\mu} = 0$$

$$2\sum_{ij}\left(y_{ij} - \mu - t_i\right)(-1) = 0$$

$$-2\sum_{ij}\left(y_{ij} - \mu - t_i\right) = 0$$

$$\sum_{ij}\left(y_{ij} - \mu - t_i\right) = \frac{0}{-2} = 0$$

Where $\sum_i y_{ij} = G$ , G= Grand total

$$G = \sum_i n_i \mu + \sum_i n_i t_i \ \ldots (1)$$

$$\frac{\partial (e'e)}{\partial t_i} = 0$$

$$2\sum_j \left(y_{ij} - \mu - t_i\right)(-1) = 0 \quad -2\sum_j \left(y_{ij} - \mu - t_i\right) = 0$$

$$\sum_j \left(y_{ij} - \mu - t_i\right) = \frac{0}{-2} = 0$$

$$\sum_j y_{ij} - \sum_j \mu - \sum_j t_i = 0$$

Where $\sum_j y_{ij} = T_i$

$$T_i = n_i \mu + n_i t_i = 0 \ \ldots (2)$$

From equation (1)

$$\sum_i n_i t_i = 0, \qquad \sum_i n_i = n$$

$$G = n\hat{\mu} + 0$$

$$\frac{G}{n} = \hat{\mu}$$

From equation (2)

$$T_i - n_i \hat{\mu} = n_i t_i$$

$$\frac{T_i}{n_i} - \frac{n_i G}{n_i n} = \hat{t}_i$$

$$\frac{T_i}{n_i} - \frac{G}{n} = \hat{t}_i$$

Error Sum of Squares

$$E(e'e)^2 = \sum_{ij}\left(y_{ij} - \mu - t_i\right)^2$$

$$\sum_{ij}\left(y_{ij} - \mu - t_i\right)\left(y_{ij} - \mu - t_i\right)$$

$$\sum_{ij} y_{ij}\left(y_{ij} - \mu - t_i\right) + \textit{other terms are vanished}$$

$$\sum_{ij}\left[y_{ij}^2 - \hat{\mu}\, y_{ij} - \hat{t}_i\, y_{ij}\right]$$

$$= \sum_{ij} y_{ij}^2 - \hat{\mu}\sum_{ij} y_{ij} - \hat{t}_i \sum_{ij} y_{ij}$$

$$= \sum_{ij} y_{ij}^2 - \hat{\mu}\sum_{ij} y_{ij} - \sum_{i}\hat{t}_i \sum_{j} y_{ij}$$

$$= \sum_{ij} y_{ij}^2 - \frac{G}{n}G - \sum_{i}\left(\frac{T_i}{n_i} - \frac{T_i G}{n}\right)$$

$$= (\sum_{ij} y_{ij}^2 - \frac{G^2}{n} - \left(\sum_{i}\frac{T_i^2}{n_i} - \frac{G^2}{n}\right)$$

Where $\sum_{j} y_{ij} = T_i$

Error Sum of Square (E.S.S) = Total Sum of Square (T.S.S) – Treatment Sum of Square (Tr. S.S)

Where $\frac{G^2}{n}$ is the correction factor

**Table 2.1 : Anova Table for CRD**

| Source of variation | d.f | Sum of Square(SS) | Mean Sum of Square(MSS) | F-ratio |
|---|---|---|---|---|
| Treatments | k-1 | Tr.S.S= $\sum_i \frac{T_i^2}{n_i} - \frac{G^2}{n}$ | MSST= $\frac{Tr.S.S}{k-1}$ | $F = \frac{MSST}{MSSE}$ |

| | | | | |
|---|---|---|---|---|
| Error | n-k | By subtraction<br>E.S.S=T.S.S-Tr.S.S | MSSE= $\frac{Er.S.S}{n-k}$ | |
| Total | n-1 | $\sum_{ij} y_{ij}^2 - \frac{G^2}{n}$ | | |

Under the null hypothesis, $H_0 = t_1 = t_2 = \cdots = t_k$ against the alternative that all t's are not equal, the test statistic $F = \frac{MSST}{MSST} \sim F_{(k-1, n-k)}$

i.e., F follows F (central) distribution with (k-1, n-k) d.f

If $F > F_{(k-1, n-k)}$ (α) then $H_0$ is rejected at α % level of significance and we conclude that treatments differ significantly. Otherwise $H_0$ accepted.

**Problem 2.1 :**

A set of data involving four "tropical feed stuffs A, B, C, D" tried on 20 chicks is given below. All the twenty chicks are treated alike in all respects except the feeding treatments and each feeding treatment is given to 5 chicks. Analyse the data.

| Feed | Gain in Weight | Total $T_i$ |
|---|---|---|
| A | 55 49 42 21 52 | 219 |
| B | 61 112 30 89 63 | 355 |
| C | 42 97 81 95 92 | 407 |
| D | 169 137169 85 154 | 714 |
| | Grand Total | G = 1,695 |

Figures in antique in the Table are not given in the original data. They are a part of the calculations for analysis.

Weight gain of baby chicks fed on different feeding materials composed of tropical feed stuffs is given in Table.

Solution:

Null hypothesis, $H_o$: $t_A = t_B = t_C = t_D$

i.e., the treatment effects are same. In other words, all the treatments (A, B, C, D) are alike as regards their effect on increase in weight.

Alternative hypothesis, $H_1$: At least two of $t_i$'s are different.

Raw S.S. (R.S.S.) = $\sum_i \sum_j y_{ij}^2$ = $55^2 + 49^2 + \ldots\ldots\ldots + 85^2 + 154^2$ = 1, 81,445

Correction factor (C.F.) = $G^2/N = (1,695)^2/20$ = 1, 43,651.25

Total S.S. (T.S.S.) = R.S.S. – C.F. = 1, 81.445 – 1, 43,651.25 = 37, 793.75

Treatment S.S. = $T_1^2 + T_2^2 + T_3^2 + T_4^2/5$ – C.F.

= 47, 961 + 1, 26,025 + 1, 65, 649 + 5, 09,769/5 – 1, 43,641.25

Error S.S. = Total S.S. – Treatment S.S. = 37,793.75 – 26,234.95 = 11,558.80

Table 3.2 : Anova Table for CRD

| Source of variation | S.S. | d.f. | M.S.S. = S.S./d.f. | Variance ratio, 'F' |
|---|---|---|---|---|
| Treatments | 26,234.95 | 3 | 8744.98 | FT = 8744.98/722.42 = 12.105 |
| Error | 11,558.80 | 16 | 722.42 | |
| Total | 37,793.75 | 19 | | |

Test statistic: FT ~ F(3,16), Tabulated $F_{o.o5}$ (3, 16) = 3.06. Hence FT is highly significant and we rejected $H_o$ at 5% level of significance and conclude that the treatments A, B, C and D differ significantly.

## 2.4 Randomised Block Design(RBD)

If all the treatments are applied at random relatively homogeneous units within each strata or block and replicated over all the blocks. The design is a randomised block design.

### Advantages of RBD

(i) Accuracy:

This design has been shown to be more efficient or accurate than C.R.D for most types of experimental work. The elimination of between S.S. from residual S.S. usually results in a decrease of error mean S.S.

(ii) Flexibility:

In R.B.D no restriction are placed on the number of treatments or the number of replicates. In general, at least two replicates are required to carry out the test of significance (factorial design is an exception). In addition, control (check) or some other treatments may be included more than once without complications in the analysis.

(iii) Ease of Analysis:

Statistical analysis is simple and rapid. More-over the error of any treatment can be isolated and any number of treatments may be omitted from the analysis without complicating it.

**Disadvantages of RBD**

i) RBD may give misleading results if blocks are not homogeneous.
ii) RBD is not suitable for large number of treatments in that case the block size will increase and it may not be possible to keep large blocks homogeneous.
iii) If the data on more than two plots is missing, the statistical analysis becomes quite tedious and complicated.

Layout of RBD: -

Let us consider five Treatments A, B, C, D, E each replicated 4 times we divided the whole experimental area into 4 relatively homogeneous block and each in to 5 units the treatments allocated at random to the blocks particular layout may be follows.

| | | | | | |
|---|---|---|---|---|---|
| *BlockI* | *A* | *B* | *C* | *D* | *E* |
| *BlockII* | *B* | *C* | *D* | *E* | *A* |
| *BlockIII* | *C* | *D* | *E* | *A* | *B* |
| *BlockIV* | *D* | *E* | *A* | *B* | *C* |

Lay out:

| | | | | | means | total |
|---|---|---|---|---|---|---|
| | $y_{11}$ | $y_{12}$ | .... | $y_{1r}$ | $\bar{y}_{1.}$ | $T_{1.}$ |
| | $y_{21}$ | $y_{22}$ | .... | $y_{2r}$ | $\bar{y}_{2.}$ | $T_{2.}$ |
| | .. | ... | ... | ... | ... | .. |
| | $y_{i1}$ | $y_{i2}$ | .... | $y_{ij}y_{ir}$ | $\bar{y}_{i.}$ | $T_{i.}$ |
| | ... | ... | ... | ... | .... | .... |
| | ... | ... | ... | .... | .... | ..... |
| | $y_{t1}$ | $y_{t2}$ | .... | $y_{tr}$ | $\bar{y}_{t.}$ | $T_{t.}$ |
| means | $\bar{y}_{.1}$ | $\bar{y}_{.2}$ | .... | $\bar{y}_{.r}$ | | $\downarrow$ |
| total | $T_{.1}$ | $T_{.2}$ | .... | $T_{.r}$ | $\rightarrow$ | $G$ |

Let us assume that $y_{ij}$ is the response of the yield of experiment unit from $i^{th}$ treatment $j^{th}$ block.

### 2.4.1 Statistical Analysis of RBD

The model is

$$y_{ij} = \mu + t_i + b_j + e_{ij} \; ; \; (i = 1,2,\ldots,t; \quad j = 1,2,\ldots,r)$$

Where $y_{ij}$ is the response or the yield of the experimental unit receiving the $i^{th}$ treatment in the $j^{th}$ block;

$\mu$ is the general mean effect

$t_i$ is the effect due to the $i^{th}$ treatment

$b_j$ is the effect due to $j^{th}$ block or replicate

$$e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$$

Where $\mu$ , $t_i$ and $b_j$ are constants so that $\sum_{i=1}^{t} t_i = 0 \quad and \quad \sum_{j=1}^{r} b_j = 0$

If we write $\sum_i \sum_j y_{ij} = \; G = Grand\ Total$

$\sum_j y_{ij} = T_i$ = Total for $i^{th}$ treatment

$\sum_i y_{ij} = B_j$ = Total for $j^{th}$ block

μ , $t_i$ and $b_j$are estimated by the method of least squares

$$E = \sum_i \sum_j e_{ij}^2 = \sum_i \sum_j \left(y_{ij} - \mu - t_i - b_j\right)^2 \quad \ldots (1)$$

Differentiate with respect to μ

$$\frac{\partial E}{\partial \mu} = 0$$

$$2\sum_i \sum_j (y_{ij} - \mu - t_i - b_j)(-1) = 0$$

$$-2\sum_i \sum_j (y_{ij} - \mu - t_i - b_j) = 0$$

$$\sum_i \sum_j (y_{ij} - \mu - t_i - b_j) = \frac{0}{-2} = 0$$

$$\sum_i \sum_j y_{ij} - \sum_i \sum_j \mu - \sum_i \sum_j t_{i\ i} - \sum_i \sum_j b_j) = 0$$

$$\sum_i \sum_j y_{ij} - tr\mu - r\sum_i t_i - t\sum_j b_j = 0$$

Where $\sum_i \sum_j y_{ij} = G$

$$G - tr\mu - r\sum_i t_i - t\sum_j b_j = 0 \ \ldots (2)$$

Differentiate with respect to $t_i$

$$\frac{\partial E}{\partial t_i} = 0$$

$$2\sum_j (y_{ij} - \mu - t_i - b_j)(-1) = 0$$

$$-2\sum_j (y_{ij} - \mu - t_i - b_j) = 0$$

$$\sum_j (y_{ij} - \mu - t_i - b_j) = \frac{0}{-2} = 0$$

$$\sum_j y_{ij} - \sum_j \mu - \sum_j t_i - \sum_j b_j) = 0$$

Where $\sum_j y_{ij} = T_i$

$$T_i = r\mu + rt_i - \sum_j b_j \quad \ldots(3)$$

Differentiate with respect to $b_j$

$$\frac{\partial E}{\partial b_j} = 0$$

$$2\sum_i (y_{ij} - \mu - t_i - b_j)(-1) = 0$$

$$-2\sum_i (y_{ij} - \mu - t_i - b_j) = 0$$

$$\sum_i (y_{ij} - \mu - t_i - b_j) = \frac{0}{-2} = 0$$

$$\sum_i y_{ij} - \sum_i \mu - \sum_i t_i - \sum_i b_j) = 0$$

Where $\sum_i y_{ij} = B_j$

$$B_j = t\mu + \sum_i t_i + tb_j \quad \ldots.(4)$$

$$\sum_{i=1}^{t} t_i = 0 \quad and \quad \sum_{j=1}^{r} b_j = 0$$

From equation (2)

$$G = tr\mu$$

$$\frac{G}{tr} = \hat{\mu}$$

From Equation (3)

$$T_i = r\hat{\mu} + rt_i$$

$$T_i - r\hat{\mu} = rt_i$$

$$\frac{T_i}{r} - \frac{G}{tr} = \hat{t}_i$$

From equation (4)

$$B_j - t\hat{\mu} = tb_j$$

$$\frac{B_j}{t} - \frac{G}{tr} = \hat{b}_j$$

Error Sum of Square

$$E = \sum_i \sum_j \left(y_{ij} - \mu - t_i - b_j\right)^2$$

$$E = \sum_i \sum_j (y_{ij} - \mu - t_i - b_j)(y_{ij} - \mu - t_i - b_j)$$

$$E = \sum_i \sum_j y_{ij}\,(y_{ij} - \mu - t_i - b_j) + \text{other terms are vanished}$$

$$\text{E}= \sum_i \sum_j y_{ij}^2 - \hat{\mu} \sum_i \sum_j y_{ij} - \sum_i \sum_j y_{ij}\hat{t}_i - \sum_i \sum_j y_{ij}\hat{b}_j$$

$$= \sum_i \sum_j y_{ij}^2 - \frac{G}{tr}G - \sum_i \sum_j y_{ij}\left(\frac{T_i}{r} - \frac{G}{tr}\right) - \sum_i \sum_j y_{ij}\left(\frac{B_j}{t} - \frac{G}{tr}\right)$$

Where $\sum_i \sum_j y_{ij} = G;\ \sum_j y_{ij} = T_i;\ \sum_i y_{ij} = B_j$

$$= \left(\sum_i \sum_j y_{ij}^2 - \frac{G^2}{tr}\right) - \left(\frac{\sum_i T_i^2}{r} - \frac{G^2}{tr}\right) - \left(\frac{\sum_j B_j^2}{t} - \frac{G^2}{tr}\right)$$

Error Sum of Square (E.S.S) = Total Sum of Square (T.S.S)-Treatment Sum of Square (Tr. S.S)- Block Sum of Square (B.S.S)

Where, correction factor= $\frac{G^2}{tr}$

Total Sum of Square = $\sum_i \sum_j y_{ij}^2 - \frac{G^2}{tr}$

Treatment Sum of Square = $ST^2 = \frac{\sum_i T_i^2}{r} - \frac{G^2}{tr}$

Block Sum of Square = $SB^2 = \frac{\sum_j B_j^2}{t} - \frac{G^2}{tr}$

**Table 2.2 ANOVA Table for RBD**

| Source of variation | Degrees of freedom | Sum of squares | Mean sum of square | Variance ratio |
|---|---|---|---|---|
| Treatment | (t-1) | $ST^2$ | $ST^2=ST^2/t-1$ | $FT=ST^2/SE^2$ |
| Blocks or replicates | (r-1) | $SB^2$ | $SB^2=SB^2/r-1$ | $FB^2=SB^2/SE^2$ |
| Error | (t-1) (r-1) | $SE^2$ | $SE^2=SE^2/(t-1)(r-1)$ | |
| Total | rt-1 | | | |

Under the null hypothesis $H_{0t} = t_1=t_2=\ldots=t_i$ against the alternative that all t's are not equal the test statistic is :

$$F_T = \frac{S_T^2}{S_E^2} \sim F_{[(t-1),(t-1)(r-1)]}$$

i.e., $F_T$ follows F(central) distribution with [(t-1), (t-1)(r-1)] d.f. Thus if $F_T$ is greater than tabulated F for [(t-1), (t-1)(r-1)] d.f, at certain level of significance, usually 5 % then we reject the null hypothesis$H_{0t}$ and conclude that the treatments differ significantly. If $F_t$ is less than tabulated value then $F_T$ is not significant and we conclude that the data do not provide any evidence against the null hypothesis which may be accepted.

Similarly under the null hypothesis $H_{ob}=b_1=b_2=\ldots=b_r$, against the alternative that b' s are not equal, the test statistics is:

$$F_T = \frac{S_T^2}{S_E^2} \sim F_{[(t-1),(t-1)(r-1)]}$$

And we discuss its significance as explained above.

**Problem 3.3**

Consider the results given in the following table for an experiment involving six treatments in four randomized blocks. The treatments are indicated by numbers within parentheses.

**Table 2.3**

| Blocks | Yield for a randomized block experiment treatment and yield |
|---|---|
| 1 | 24.7 (1) 27.7(3) 20.6(2) 16.2(4) 16.2(5) 24.9(6) |
| 2 | 22.7(3) 28.8(2) 27.3(1) 15.0(4) 22.5(6) 17.0(5) |
| 3 | 26.3(6) 19.6(4) 38.5(1) 36.8(3) 39.5(2) 15.4(5) |
| 4 | 17.7(5) 31.0(2) 28.5(1) 14.1(4) 34.9(3) 22.6(6) |

Test whether the treatments differ significantly.

Solution:

Null hypothesis:

$H_o t$ : $\tau_1 = \tau_2 = \tau_3 = \tau_4$ and $H_o b$ : $b_1 = b_2 = b_3 = b_4$,i.e.,the treatments as well as block are homogeneous.

Alternative hypothesis:

$H_1 t$: At least two $\tau_i$'s are different. ; $H_1 b$: At least two $b_i$'s are different.

For finding the various S.S., we rearrange the above table as follows:

**Table 2.4**

| Blocks | (1) | (2) | (3) | (4) | (5) | (6) | Block Total ($B_J$) | $B_J^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 24.7 | 20.6 | 27.7 | 16.2 | 16.2 | 24.9 | 130.0 | 16,900.00 |
| 2 | 27.3 | 28.8 | 22.9 | 15.0 | 17.0 | 22.5 | 133.3 | 17,768.89 |
| 3 | 38.5 | 39.5 | 36.8 | 19.6 | 15.4 | 26.3 | 176.1 | 31,011.21 |
| 4 | 28.5 | 31.0 | 34.9 | 14.1 | 17.7 | 22.6 | 148.8 | 22,141.44 |
| Treatment totals ($T_i$) | 199.0 | 119.9 | 122.1 | 64.9 | 66.3 | 96.3 | 388.5=G | |
| $T_i^2$ | 14,161.00 | 14,376.01 | 14,908.41 | 4,212.01 | 4,395.69 | 9,273.69 | | |
| Average | 29.75 | 30.0 | 30.5 | 16.2 | 16.6 | 24.1 | | |

Correction Factor = (3, 46,332.25/24) = 14,430.51

Raw S.S = $\sum_i \sum_j y_{ij}^2$ = 15,780.76

Total S.S = R.S.S. – C.F. = 15,780.76 – 14,430.51 = 1,350.25

S.S. due to treatments (S.S.T) = ¼ ∑ $T_i^2$ - C.F = (61,326.81/4)- 14,430.51 = 901.19

S.S due to blocks (S.S.B) = 1/6 ∑ $B_J^2$ - C.F = 87,899.63/6 – 14430.51 = 219.43

Error S.S = T.S.S. – S.S.T. – S.S.B. = 1,350.25 – 901.19 – 219.43 = 229.63.

**Table 2.5: Anova Table**

| Source of variation | d.f. | S.S. | M.S.S. | Variance ratio (F) |
|---|---|---|---|---|
| Treatment | 5 | 901.19 | $s^2T$ = 180.24 | Ft = 180.24/15.31 = 11.8 |

| | | | | |
|---|---|---|---|---|
| Block<br>Error | 3<br>15 | 219.43<br>229.63 | $s^2B = 73.14$<br>$s^2E = 15.31$ | Fb = 73.14/15.31 = 4.7 |
| Total | 23 | 1.350.25 | | |

Tabulated $F_{3, 15}$, (0.05) = 5.42 and $F_{5, 15}$ (0.05) = 4.5 .Since under $H_o$t Ft ~ F (5, 15) and under $H_o$b, Fb ~ F (3, 15), we see that Ft is significant while Fb is not significant at 5% level of significance. Hence, Ft is rejected at 5% level of significance and we conclude that treatment effects are not alike. On the other hand, Hb may be retained at 5% level of significance and we may conclude that the blocks are homogeneous.

### 2.4.2 Estimation of one Missing Value in RBD

Let the observation $y_{ij}$ = x (say) in the $j^{th}$ block and receiving the $i^{th}$ treatment be missing, as given in table 3.7 **Table 3.7**

| | | | | | Treatments | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | … | I | … | t | |
| Blocks | 1 | $Y_{11}$ | $Y_{21}$ | … | $Y_{i1}$ | … | $Y_{t1}$ | $y_{.1}$ |
| | 2 | $Y_{12}$ | $Y_{22}$ | … | $Y_{i2}$ | … | $Y_{t2}$ | $y_{.2}$ |
| | ⋮ | ⋮ | ⋮ | … | ⋮ | … | ⋮ | ⋮ |
| | J | $Y_{j1}$ | $Y_{j2}$ | … | X | … | $y_{jt}$ | $y_{.j}' + x$ |
| | ⋮ | ⋮ | ⋮ | … | ⋮ | … | ⋮ | ⋮ |
| | R | $Y_{1r}$ | $Y_{2r}$ | … | $y_{ir}$ | … | $y_{tr}$ | |
| | Total | $Y_{1.}$ | $Y_{2.}$ | … | $y_{i.}' + x$ | … | $Y_{t.}$ | $y_{..}' + x$ |

where

$y_{i.}'$ is total of known observations getting $i^{th}$ treatment

$y_{.j}'$ is total of known observations in $j^{th}$ block and

$y..'$ is total of all known observations

Correction factor = $\frac{G^2}{tr} = \frac{(G'+x)^2}{tr}$

Total sum of square = $\sum_i \sum_j y_{ij}^2 - \frac{G^2}{tr} = x^2 + cons\tan t\ terms\ independent\ of\ x - \frac{(G'+x)^2}{tr}$

Sum of square due to treatment (S.S.Tr) = $\frac{\sum_i y_i.'}{r} - C.F = \frac{\sum_i \left(y_{i..}' + x\right)^2}{r} - C.F$

Sum of square due to

Block (S.S.B) = $\frac{\sum_j y_{.j}^{\ 2}}{t} - C.F = \frac{\left(y_{.j}' + x\right)^2}{r} - C.F$

Sum of square due to error = T.S.S –S.S.Tr-S.S.B =

$$(\sum_i \sum_j y_{ij}^2 - C.F) - \left(\frac{\sum_i y_{i.}^2}{r} - C.F\right) - (\frac{\sum_j y_{.j}^2}{t} - C.F)$$

=

$$x^2 + cons\tan t\ terms\ \ independent\ of\ \ x - \frac{(G'+x)^2}{tr} - \left[\frac{\left(y_{i.'}' + x\right)^2}{r} - \frac{(G'+x)^2}{tr}\right] - \left[\frac{(y.j'+x)^2}{t} - \frac{(G'+x)^2}{tr}\right]$$

$$= x^2 + cons\tan t\ terms\ independent\ \ of\ \ x - \frac{(G'+x)^2}{tr} - \frac{\left(y_{i.}' + x\right)^2}{r} + \frac{(G'+x)^2}{tr} - \frac{(y.j'+x)^2}{t} + \frac{(G'+x)^2}{tr}$$

$$= x^2 + cons\tan t\ terms\ independent\ of\ \ x - \frac{\left(y_{i.}' + x\right)^2}{r} - \frac{(y.j'+x)^2}{t} + \frac{(G'+x)^2}{tr}$$

Differentiate with respect to x

$$\frac{\partial(S.S.E)}{\partial x} = 0$$

$$= 2x - 2\frac{(yi.' + x)}{r} - \frac{2(y.j' + x)}{t} + 2\frac{(G' + x)}{tr} = \frac{0}{2} = 0$$

$$x - \frac{\left(y_{i.}^{'} + x\right)}{r} - \frac{\left(y_{ij}^{'} + x\right)}{t} + \frac{(G' + x)}{tr} = 0$$

$$\frac{trx - t(y_{i.}^{'} + x) - r\left(y_{.j}^{'} + x\right) + (G' + x)}{tr} = 0$$

$$trx - t(yi.' + x) - r(y.j' + x) + (G' + x) = 0 \times tx = 0$$

$$trx - ty_{i.}^{'} + tx - ry_{.j}^{'} + rx + (G' + x) = 0$$

$$x(tr - t - r + 1) - ty_{i.}^{'} - ry_{.j}^{'} + G' = 0$$

$$x(tr - t - r + 1) = ty_{i.}^{'} + ry_{.j}^{'} - G'$$

$$x((t-1)(r-1)) = ty_{i.}^{'} + ry_{.j}^{'} - G'$$

$$x = \frac{ty_{i.}^{'} + ry_{.j}^{'} - G'}{(r-1)(t-1)}$$

**Problem 3.3**

Suppose that the value for treatment 2 is missing in replication III. The data will then be as presented in the table below.

**Table 2.6 RBD data with one missing value.**

| Treatment | Replication | | | | Total |
|---|---|---|---|---|---|
| | I | II | III | IV | |
| 1 | 22.9 | 25.9 | 39.1 | 33.9 | 121.8 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 29.5 | 30.4 | X | 29.6 | 89.5 |
| 3 | 28.8 | 24.4 | 32.1 | 28.6 | 113.9 |
| 4 | 47.0 | 40.9 | 42.8 | 32.1 | 162.8 |
| 5 | 28.9 | 20.4 | 21.1 | 31.8 | 102.2 |
| Total | 157.1 | 142.0 | 135.1 | 156.0 | 590.2 |

$X = rR' + tT' - G'/ (r-1)(t-1)$

$= 4(135.1) + 5(89.5) - 590.2/(3)(4)$

$= 397.7/12$

$= 33.1$

The upward bias,

$B = [R' - (t-1) X]^2/t(t-1)$

$= [135.1 - 4(33.1)]^2/(5)(4)$

$= 7.29/20$

$= 0.3645$

After substituting the estimated missing value, we get

Treatment 2 total = 89.5 + 33.1 = 122.6,

Replication 3 total = 135.1 + 33.1 = 168.2, and

The grand total = 590.2 + 33.1 = 623.3

Treatment SS = ¼ $[(121.8)^2 + (122.6)^2 + (113.9)^2 + (162.8)^2 + 102.2)^2] - (623.3)^2/20$

$= 19946.9725 - 19425.1445$

$= 521.8280$

Corrected treatment SS = 521.8280 – 0.3645

= 521.4635

With these values the analysis of variance table is completed.

**Table 2.7 Analysis of variance for the data in Table**

| Source of variation | df | SS | MS | F |
|---|---|---|---|---|
| Replication | 3 | 69.1855 | 23.0618 | 1 |
| Treatment | 4 | 521.4635 | 130.3659 | 4.117 |
| Error | 11 | 347.9475 | 31.6316 | |
| Total | 18 | 938.9610 | | |

### 2.4.3 Estimation of two missing values

Suppose in RBD with k treatments and R-Replications, two observations are missing. Let x and y be two missing observations and they belong two different Block and affected different treatment. We assume that x belongs to the jth to the ith treatments and y belong to ith block and $m^{th}$ treatment. Estimate the missing observations x and y.

**Layout of two missing observations in RBD.**

| | 1…… | 2……. | 1…….m…… | K | |
|---|---|---|---|---|---|
| **1** | $y_{11}$ | $y_{12}$ | | | **$B_1$** |
| **2** | $y_{21}$ | $y_{22}$ | | | **$B_2$** |
| . | | | | | . |
| . | | | | | . |
| **J** | | **x** | | | **$B'_J$+x** |

.  .

.  .

**I  y  $B'_i+y$**

.  .

.  .

**r  $B_r$**

**$T_1$  $T_2$  $T_i'+X$.....  $T'_m+Y$  $T_k$  $G'+x+y$**

Correction factor $= \frac{G^2}{tr} = \frac{(G'+x+y)^2}{tr}$

Total sum of square =

$$\sum_i \sum_j y_{ij}^2 - \frac{G^2}{tr} = x^2 + y^2 + cons\tan t\ terms\ \ independent\ of\ \ x\ and\ y - \frac{(G'+x+y)^2}{tr}$$

Sum of square due to treatment (S.S.Tr) = $\frac{\sum_i y_i.'}{r} - C.F = \frac{\sum_i \left( y_{i..}{}' + x + y \right)^2}{r} - C.F$

Sum of square due to

Block (S.S.B) = $\frac{\sum_j y_{.j}{}^2}{t} - C.F = \frac{\left( y_{.j}{}' + x + y \right)^2}{r} - C.F$

S.S.E=T.S.S – S.S.Tr-S.S.B

$$\left[ x^2 + y^2 + cons\tan t\ terms\ indepent\ of\ \ x\ and\ y - C.F \right]$$

$$= -\left[ \left( \frac{T_i{}' + x}{r} \right)^2 + \left( \frac{T_m{}' + y}{r} \right)^2 - C.F \right] - \left[ \left( \frac{B_j{}' + x}{t} \right)^2 + \left( \frac{B_i{}' + y}{t} \right)^2 - C.F \right]$$

$$= \begin{bmatrix} x^2 + y^2 + cons\tan t\ terms\ indepent\ of\ x\ and\ y - .C.F - \dfrac{\left(T_i^{'} + x\right)^2}{r} - \dfrac{\left(T_m^{'} + y\right)^2}{r} \\ C.F - \dfrac{\left(B_j^{'} + x\right)^2}{t} - \dfrac{\left(B_i^{'} + y\right)^2}{t} + C.F \end{bmatrix}$$

$$= x^2 + y^2 - \frac{\left(T_i^{'} + x\right)^2}{r} - \frac{\left(T_m^{'} + y\right)^2}{r} - \frac{\left(B_j^{'} + x\right)^2}{t} - \frac{\left(B_i^{'} + y\right)^2}{t} + \frac{(G' + x + y)^2}{tr} \quad \ldots(1)$$

Differentiate with respect to x in equation (1)

$$\frac{\partial.S.S.E}{\partial x} = 0$$

$$2x - \frac{2(T_i^{'} + x)}{r} - \frac{2\left(B_j^{'} + x\right)}{t} + \frac{2(G' + x + y)}{tr} = 0$$

$$x - \frac{(T_i^{'} + x)}{r} - \frac{\left(B_j^{'} + x\right)}{t} + \frac{(G' + x + y)}{tr} = \frac{0}{2} = 0$$

$$\frac{xtr - t\left(T_i^{'} + x\right) - r\left(B_j^{'} + x\right) + (G' + x + y)}{tr} = 0$$

$$xtr - t\left(T_i^{'} + x\right) - r(B_j^{'} + x) + (G' + x + y) = 0 \times tr = 0$$

$$xtr - tT_i^{'} - tx - rB_j^{'} - rx + G' + x + y = 0$$

$$x(tr - t - r + 1) = tT_i^{'} + rB_j^{'} - G' - y$$

$$x = \frac{tT_i^{'} + rB_j^{'} - G' - y}{(t-1)(r-1)}$$

Differentiate with respect to y in equation (1)

$$\frac{\partial.S.S.E}{\partial y} = 0$$

$$2y - \frac{2(T_m^{'} + y)}{r} - \frac{2\left(B_i^{'} + y\right)}{t} + \frac{2(G' + x + y)}{tr} = 0$$

$$y - \frac{(T_m^{'} + y)}{r} - \frac{\left(B_i^{'} + y\right)}{t} + \frac{(G' + x + y)}{tr} = \frac{0}{2} = 0$$

$$\frac{ytr - t\left(T_m^{'} + y\right) - r\left(B_i^{'} + y\right) + \left(G' + x + y\right)}{tr} = 0$$

$$ytr - t\left(T_m^{'} + y\right) - r(B_i^{'} + y) + \left(G' + x + y\right) = 0 \times tr = 0$$

$$ytr - tT_m^{'} - ty - rB_i^{'} - ry + G' + x + y = 0$$

$$y(tr - t - r + 1) = tT_m^{'} + rB_i^{'} - G' - x$$

$$y = \frac{tT_m^{'} + rB_i^{'} - G' - x}{(t-1)(r-1)}$$

**Problem 3.4**

Suppose that one more value is missing in row 5 and column 3.

Table 2.8 Grain yield of paddy, kg/plot

| E | C | D | B | A | Total |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| 26 | 42 | 39 | 37 | 24 | 168 |
| A<br>24 | D<br>33 | E<br>21 | C<br>(X) | B<br>38 | 116 |
| D<br>47 | B<br>45 | A<br>31 | E<br>29 | C<br>31 | 183 |
| B<br>38 | A<br>24 | C<br>36 | D<br>41 | E<br>34 | 173 |
| C<br>41 | E<br>24 | B<br>(X) | A<br>26 | D<br>30 | 121 |
| Total | 68 | 127 | 133 | 157 | 761 |

The treatment totals are

A : 129, B : 158, C : 150, D : 190,E : 134

The means for second row and fourth column in which C is missing are 116/4 = 29.0 and 133/4 = 33.25, respectively. Hence the first estimate for C is

$C_1 = 29.00 + 33.25/2 = 31.12$

G' = 761 + 31.12 = 792.12

$B_1 = t(R' + C' + T') - 2G'/(t-1)\ (t-2)$

$= 5(121 + 127 + 158) - 2(792.12)/(5-1)(5-2)$

$= 2030/12 - 1584.24/12$

$= 169.17 - 132.02$

$= 37.15$

For the second cycle we have

G' = 761 + 37.15 = 798.15

$G_2 = 5(116 + 133 + 150) – 2(798.15)/12$

$= 1995/12 – 1596.3/12$

$= 166.25 – 133.03$

$= 33.22$

$G' = 761 + 33.22$

$= 794.22$

$B^2 = 169.17 – 2(794.3)/12$

$= 36.8$

It can be seen that the estimated values for B are same and that for C are very close. Hence we stop the iteration process at third cycle. The final estimates for B and C for the missing plots are 36.8 and 33.3 respectively.

The column total, row total, etc., with respect to the missing plots are modified by adding the estimated values. Thus we have,

Treatment B total = 158 + 36.8 = 194.8

Treatment C total = 150 + 33.3 = 183.3

Second row total = 116 + 33.3 = 149.3

Fifth row total = 121 + 36.8 = 157.8

Third column total = 127 + 36.8 = 163.8

Fourth column total = 133 + 33.3 = 166.3

Grand total = 761 + 36.8 + 33.3 = 831.1

The data is then analysed in the usual manner.

$CF = (831.1)^2/25$

$= 27629.0884$

Total SS = 28902.130 – CF =1273.0416

Row SS = 27766.666 – CF = 137.5776

Column SS = 27667.026 – CF = 37.9376

Treatment SS = 28448.586 – CF = 819.4976

Error SS = 278.0288

Now ignoring the treatment classification the missing values are estimated as in the case of RBD. The estimate of the second row, fourth column missing value is 28.5; and that of fifth row, third column is 28.2. After substituting the estimated values and analyzing the data as RBD, we get the error sum of squares as 1031.5856. Then we have,

Corrected treatments SS = Error SS (RBD) – Error (LSD)

= 1031.5856 – 278.0288

= 753.5568

The final results are presented in the following table.

Table 2.9: Analysis of variance for the data

| Source of variation | df | SS | MS | F |
|---|---|---|---|---|
| Row | 4 | 137.5776 | 34.3944 | 1.237 |
| Column | 4 | 37.9376 | 9.4844 | <1 |
| Treatment | 4 | 753.5568 | 188.3892 | 6.776 |
| Error | 10 | 278.0288 | 27.8029 | |
| Total | 22 | 1273.0416 | | |

## 2.5 Latin square design (LSD)

LSD is defined for eliminating the variation of two factors called row and column in this design. The number of treatments is equal to the number of replications.

Layout of design

In this design the number of treatments is equal to the number of replications. Thus in case of m treatments there have to be mxm = $m^2$ experimental units. The whole of the experimental area is divided into $m^2$

experimental units (plots) arranged in a square so that each row as well each column contain m units.

The m treatments are allocated at random to these rows and columns in such a way that every treatment occurs only once in each row and in each column. Such a layout is LSD.

**2x2 layouts**

| | |
|---|---|
| **A** | **B** |
| **B** | **A** |

**3x3 layouts**

| | | |
|---|---|---|
| **A** | **B** | **C** |
| **B** | **C** | **A** |
| **C** | **A** | **B** |

**4x4 layouts**

| | | | |
|---|---|---|---|
| **A** | **B** | **C** | **D** |
| **B** | **C** | **D** | **A** |
| **C** | **D** | **A** | **B** |
| **D** | **A** | **B** | **C** |

**5X5 layouts**

| | | | | |
|---|---|---|---|---|
| **A** | **B** | **C** | **D** | **E** |
| **B** | **C** | **D** | **E** | **A** |
| **C** | **D** | **E** | **A** | **B** |

| | | | | |
|---|---|---|---|---|
| **D** | **E** | **A** | **B** | **C** |
| **E** | **A** | **B** | **C** | **D** |

**Example:**

An animal feeding experiment where the column groups may corresponding with initial weight and the row group with age.

**Standard Latin square:**

A Latin in which the treatments say A, B, C etc occur in the first row and first column in alphabetical order is called standard Latin square.

**Example:**

| | |
|---|---|
| **A** | **B** |
| **B** | **A** |

**Advantages of LSD**

1. With two way grouping LSD controls more of the variation than CRD or RBD.
2. The two way elimination of variation as a result of cross grouping often results in small error mean sum of squares.
3. LSD is an incomplete 3-way layout. Its advantage over the complete 3-way layout is that instead of $m^3$ experimental units only $m^2$ units are needed. Thus, a 4x4 LSD results in saving of $m^3 = 4^3 - 4^2 = 64-16 = 48$ observations over a complete 3-way layout.
4. The statistical analysis is simple though slightly complicated than for RBD. Even 1 or 2 missing observations the analysis remains relatively simple.
5. More than one factor can be investigated simultaneously.

**Disadvantages of LSD**

1. LSD is suitable for the number of treatments between 5 and 10 and for more than 10 to 12 treatments the design is seldom used. Since

in that case, the square becomes too large and does not remain homogeneous.

2. In case of missing plots the statistical analysis becomes quite complex.

3. If one or two blocks in a field are affected by some disease or pest. We can't omit because the number of rows columns and treatments have to be equal.

### 2.5.1 Statistical Analysis of LSD

**L**et $y_{ijk}$ (i, j, k=1,2,…,m)denote the response from the unit in the $i^{th}$ row, $j^{th}$ column and receiving the $k^{th}$ treatment.

The model is

$$y_{ijk} = \mu + r_i + c_j + t_k + e_{ijk} \; ; \qquad i, j, k = 1, 2, \ldots, m$$

Where μ is the constant mean effect; $r_i$, $c_j$ and $t_k$ due to the $i^{th}$ row, $j^{th}$ column and $k^{th}$ treatment respectively and $e_{ijk}$ is error effect due to random component assumed to be normally distributed with mean zero and variance $\sigma_e^2$ $i.e., e_{ijk} \sim N(0, \sigma_e^2)$

If we write

G= Total of all the $m^2$ observations

$R_i$ = Total of the m observations in the $i^{th}$ row

$C_j$ = Total of the m observations in the $j^{th}$ column

$T_k$ = Total of the m observations from $k^{th}$ treatment

Estimation by the method of least squares

$$E(e_{ijk})^2 = \sum_{ijk} (y_{ijk} - \mu - r_i - c_j - t_k)^2 \quad \ldots \quad (1)$$

$$\frac{\partial E}{\partial \mu} = 0, \quad \frac{\partial E}{\partial r_i} = 0, \quad \frac{\partial E}{\partial c_j} = 0, \quad \frac{\partial E}{\partial t_k} = 0$$

Differentiate with respect to μ in equation (1)

$$\frac{\partial E}{\partial \mu} = 2\sum_{ijk}\left(y_{ijk} - \mu - r_i - c_j - t_k\right)(-1) = 0$$

$$\sum_{ijk} y_{ijk} - \sum_{ijk}\mu - \sum_{ijk} r_i - \sum_{ijk} c_j - \sum_{ijk} t_k = 0$$

Where $\sum_{ijk} y_{ijk} = G$ , I,j,k=m$^2$, I,j=m, I,k=m

$$G - m^2\mu - m\sum_i r_i - m\sum_j c_j - m\sum_k t_k = 0 \ \ldots (2)$$

Differentiate with respect to $r_i$ in equation (1)

$$\frac{\partial E}{\partial r_i} = 2\sum_{jk}\left(y_{ijk} - \mu - r_i - c_j - t_k\right)(-1) = 0$$

$$\sum_{jk} y_{ijk} - \sum_{jk}\mu - \sum_{jk} r_i - \sum_{jk} c_j - \sum_{jk} t_k = 0$$

Where $\sum_{jk} y_{ijk} = R_i$ , I,j,k=m$^2$, I,j=m, I,k=m

$$R_i - m\mu - mr_i - m\sum_j c_j - m\sum_k t_k = 0 \ldots (3)$$

Differentiate with respect to $c_j$ in equation (1)

$$\frac{\partial E}{\partial c_j} = 2\sum_{ik}\left(y_{ijk} - \mu - r_i - c_j - t_k\right)(-1) = 0$$

$$\sum_{ik} y_{ijk} - \sum_{ik}\mu - \sum_{ik} r_i - \sum_{ik} c_j - \sum_{ik} t_k = 0$$

Where $\sum_{ik} y_{ijk} = C_j$ , I,j,k=m$^2$, I,j=m, I,k=m

$$C_j - m\mu - m\sum_i r_i - mc_j - m\sum_k t_k = 0 \ \ldots (4)$$

Differentiate with respect to $t_k$ in equation (1)

$$\frac{\partial E}{\partial t_k} = 2\sum_{ij}\left(y_{ijk} - \mu - r_i - c_j - t_k\right)(-1) = 0$$

$$\sum_{ij} y_{ijk} - \sum_{ij} \mu - \sum_{ij} r_i - \sum_{ij} c_j - \sum_{ij} t_k = 0$$

Where $\sum_{ij} y_{ijk} = T_k$ , I,j,k=m$^2$, I,j=m, I,k=m

$$T_k - m\mu - m\sum_i r_i - m\sum_j c_j - mt_k = 0 \ldots (5)$$

The equations (2), (3),(4) and (5) are not independent

We assume that, $\sum_i r_i = 0$ , $\sum_j c_j = 0 and \sum_k t_k = 0$

From equation (2)

$$G - m^2\mu$$

$$G = m^2\mu$$

$$\frac{G}{m^2} = \hat{\mu}$$

From equation (3)

$$R_i - m\mu - mr_i = 0$$

$$R_i - m\hat{\mu} = mr_i$$

$$\frac{R_i}{m} - \frac{mG}{m\,m^2} = \hat{r}_i$$

$$\frac{R_i}{m} - \frac{G}{m^2} = \hat{r}_i$$

From equation (4)

$$C_j - m\mu - mc_j = 0$$

$$C_j - m\hat{\mu} = mc_j$$

$$\frac{C_j}{m} - \frac{mG}{m.m^2} = \hat{c}_j$$

$$\frac{C_j}{m} - \frac{G}{m^2} = \hat{c}_j$$

From equation (5)

$$T_k - m\mu - mt_k = 0$$

$$T_k - m\hat{\mu} = mt_k$$

$$\frac{T_k}{m} - \frac{mG}{mm^2} = \hat{t}_k$$

$$\frac{T_k}{m} - \frac{G}{m^2} = \hat{t}_k$$

Error Sum of Square

$$E(e_{ijk})^2 = \sum_{ijk}(y_{ijk} - \mu - r_i - c_j - t_k)^2$$

$$= \sum_{ijk}(y_{ijk} - \mu - r_i - c_j - t_k)(y_{ijk} - \mu - r_i - c_j - t_k)$$

$$= \sum_{ijk}(y_{ijk} - \mu - r_i - c_j - t_k)(y_{ijk}) + other\ terms\ are\ vanished$$

$$= \sum_{ijk} y_{ijk}^2 - \hat{\mu}\sum_{ijk} y_{ijk} - \sum_{ijk} y_{ijk}\hat{r}_i - \sum_{ijk} y_{ijk}\hat{c}_j - \sum_{ijk} y_{ijk}\hat{t}_k$$

$$= \sum_{ijk} y_{ijk}^2 - \frac{G}{m^2}\sum_{ijk} y_{ijk} - \sum_{ijk} y_{ijk}\left(\frac{R_i}{m} - \frac{G}{m^2}\right) - \sum_{ijk} y_{ijk}\left(\frac{C_j}{m} - \frac{G}{m^2}\right) - \sum_{ijk} y_{ijk}\left(\frac{T_k}{m} - \frac{G}{m^2}\right)$$

$$= (\sum_{ijk} y_{ijk}^2 - \frac{G^2}{m^2}) - \left(\frac{\sum_i R_i^2}{m} - \frac{G^2}{m^2}\right) - \left(\frac{\sum_j C_j^2}{m} - \frac{G^2}{m^2}\right) - \left(\frac{\sum_k T_k^2}{m} - \frac{G^2}{m^2}\right)$$

Total Sum of Square $= \sum_{ijk} y_{ijk}^2 - \frac{G^2}{m^2}$

Row Sum of Square $= S_R^2 = \frac{\sum_i R_i^2}{m} - \frac{G^2}{m^2}$

Column Sum of Square= $S_C^2 = \frac{\sum_j C_j^2}{m} - \frac{G^2}{m^2}$

Treatment Sum of Square= $S_T^2 = \frac{\sum_k T_k^2}{m} - \frac{G^2}{m^2}$

Table 2.10: ANOVA Table for LSD

| Source of variation | Degrees of freedom | Sum of squares | Mean sum of square | Variance ratio |
|---|---|---|---|---|
| Rows | m-1 | $S_R^2$ | $s_R^2 = S_R^2/(m-1)$ | $F_R = s_R^2/s_E^2$ |
| Columns | m-1 | $S_C^2$ | $s_C^2 = S_C^2/(m-1)$ | $F_C = s_C^2/s_E^2$ |
| Treatments | m-1 | $S_T^2$ | $s_T^2 = S_T^2/(m-1)$ | $F_T = s_T^2/s_E^2$ |
| Error | (m-1)(m-2) | $S_E^2$ | $s_E^2 = S_E^2/(m-1)(m-2)$ | |
| Total | $m^2$-1 | | | |

Let us set up null hypothesis

For row effects $H_{0r}=r_1=r_2=\ldots=r_m=0$

For column effects $H_{0c}=c_1=c_2=\ldots=c_m=0$

For treatment effects $H_{0t}=t_1=t_2=\ldots=t_m=0$

Alternative Hypotheses

For row effects, $H_{1r}$: At least two $r_i$'s are different

For column effects, $H_{1c}$: At least two $c_i$'s are different

For treatment effects, $H_{1t}$: At least two $t_i$'s are different

d.f under the null hypotheses $H_r$, $H_b$ and $H_t$, respectively.

Let $F_\alpha = F_\alpha\{(m-1), (m-1)(m-2)\}$ be tabulated value of F for [(m-1),(m-1)(m-2)] d.f. at the level of significance $\alpha$ . Thus if $F_R > F\alpha$ we reject $H_{or}$ and if $F_R \leq F_\alpha$ we fail to reject $H_{0r}$.

Similarly, we can test for $H_{0c}$ and $H_{0t}$.

**Problem 3**

An experiment was carried out to determine the effect of claying the ground on the field of barley grains; amount of clay used were as follows:

A: No clay

B: Clay at 100 per acre

C: Clay at 200 per acre

D: Clay at 300 per acre.

The yields were in plots of 8 meters by 8 meters and are given in table.

| | I | II | III | IV | Row totals ($R_i$) |
|---|---|---|---|---|---|
| I | D<br>29.1 | B<br>18.9 | C<br>29.4 | A<br>5.7 | 83.1 |
| II | C<br>16.4 | A<br>10.2 | D<br>21.2 | B<br>19.1 | 66.9 |
| III | A<br>5.4 | D<br>38.8 | B<br>24.0 | C<br>37.0 | 105.2 |
| IV | B<br>24.9 | C<br>41.7 | A<br>9.5 | D<br>28.9 | 105.0 |
| Column Totals ($C_j$) | 75.8 | 109.6 | 84.1 | 90.7 | 306.2 |

Perform the ANOVA and calculate the critical difference for the treatment mean yields.

Solution:

The four treatment totals are:

A: 30.8, B:86.9, C:124.5, D:118.0

Grand total G = 360,2, N = 16.

C.F. = $(360.2)^2/16$ = 8109.0025

Raw S.S. = $(29.1)^2 + (18.9)^2 + \ldots\ldots\ldots + (9.5)^2 + (28.9)^2$ = 10,052.08

Total S.S. = 10,052.08 – 8,109.0025 = 1,943.0775

S.S.R. = ¼ $[(83.1)^2 + (66.9)^2 + (105.0)^2 + (105.0)^2]$ – 8,109.0025

= 33,473.26/4 – 8,109.0025 = 259.3125

S.S.C. = ¼ $[(75.8)^2 + (109.6)^2 + (84.1)^2 + (90.7)^2]$ – 8,109.0025

= 33057.10/4 – 8109.0025 = 155.2725

S.S.T. = ¼ $[(30.8)^2 + (86.9)^2 + (124.5)^2 + (118.0)^2]$ – 8,109.0025

= 37924.50/4 – 8109.0025 = 1372.1225

Error S.S. = T.S.S. – S.S.R. – S.S.C. – S.S.T. = 156.3700

ANOVA TABLE FOR L.S.D.

| Source of variation (1) | d.f. (2) | S.S. (3) | M.S.S. (4) = (3) ÷ (2) | Variance Ratio |
|---|---|---|---|---|
| Rows | 3 | 259.5375 | 86.4375 | FR = 86.4375/26.0616 = 3.32<4.76 |
| Columns | 3 | 155.2725 | 51.7575 | Fc = 51.7576/26.0616 = 1.98 <4.76 |
| Treatments | 3 | 1,372.1225 | 457.3742 | FT = 457.3742/26.0616 = 17.55 > 4.76 |
| Error | 6 | 156.3700 | 26.0616 | |
| Total | 15 | 1,943.0775 | | |

Tabulated $F_{3,\ 6}$ (0.05) = 4.76

Hence we conclude that the variation due to rows and columns is not significant but the treatments, i.e., different levels of clay, have significant effect on the yield.

### 2.5.2 One Missing observation in LSD

Let us suppose that in m×m Latin Square, the observation occurring in the $i^{th}$ row , $j^{th}$ column and receiving the $k^{th}$ treatment is missing. Let us assume that its value is x, i.e., $y_{ijk}=x$

$R_i'$ = Total of the known observations in the $i^{th}$ row.

$C_j'$ = Total of the known observations in the $j^{th}$ column.

$T_k'$ = Total of the known observations receiving $k^{th}$ treatment.

G = grand total.

$$= x^2 + cons\tan t\ terms\ independent\ of\ x - \frac{(G'+x)^2}{m^2} - \left( \frac{\left(R_i^{'} + x\right)}{m} - \frac{(G'+x)^2}{m^2} \right)$$

$$- \left( \frac{\left(C_j^{'} + x\right)^2}{m} - \frac{(G'+x)^2}{m^2} \right) - \left( \frac{\left(T_k^{'} + x\right)^2}{m} - \frac{(G'+x)^2}{m^2} \right)$$

$$= x^2 + cons\tan t\ terms\ independent\ of\ x - \frac{\left(R_i^{'} + x\right)^2}{m} - \frac{\left(C_j^{'} + x\right)^2}{m} - \frac{(T_k'+x)^2}{m} + 2\left( \frac{(G'+x)^2}{m^2} \right)$$

Differentiate w. r.to x

$$2x - \frac{2(R_i'+x)}{m} - \frac{2(C_j'+x)}{m} - \frac{2(T_k'+x)}{m} + \frac{4(G'+x)}{m^2} = 0$$

$$= x - \frac{\left(R_i^{'} + x\right)}{m} - \frac{(C_j'+x)}{m} - \frac{(T_k'+x)}{m} + \frac{2(G'+x)}{m^2} = \frac{0}{2} = 0$$

$$\frac{m^2 x}{m^2} - \frac{m\left(R_i^{'} + x\right)}{m^2} - \frac{m(C_j'+x)}{m^2} - \frac{m(T_k'+x)}{m^2} + \frac{2(G'+x)}{m^2} = 0$$

$$m^2 x - m\,|\,(R_i'+x) - m(C_j'+x) - m(T_k'+x) + 2(G'+x) = 0$$

$$m^2 x - mR_i^{'} + mx - mC_j^{'} - mx - mT_k' + 2G' + 2x = 0$$

$$m^2x - mx - mx - mx + 2x = mR_i^{'} + mC_j' + mT_k' - 2G'$$

$$x(m^2 - m - m - m + 2) = mR_i^{'} + mC_j' + mT_k' - 2G'$$

$$x((m-1)(m-2) = mR_i^{'} + mC_j' + mT_k' - 2G'$$

$$x = \frac{m(R_i' + C_j' + T_k' - 2G'}{(m-1)(m-2)}$$

Unit -III

**3.1 Post Hoc Tests in ANOVA**

Although ANOVA can be used to determine if three or more means are different, it provides no information concerning where the difference lies. For example, if Ho: $mean_1 = mean_2 = mean_3$ is rejected, then there are three alternate hypotheses that can be tested: $mean_1 \neq mean_2 \neq mean_3$, $mean_1 \neq mean_2 = mean_3$, or $mean_1 = mean_2 \neq mean_3$. Methods have been constructed to test these possibilities, and they are termed multiple comparison post-tests. There are several tests are as followed. There are,

**3.2 Multiple range test [MRT]**

In the case of significance F, the null hypothesis rejected then the problem is known which of the treatment means are significantly different. Many test procedures are available for this purpose. The most commonly used test is,

I) Lest significance difference [is known as critical difference]
II) Duncan's multiple range test [DMRT].

**3.2.1 Critical difference (C.D)**

The critical difference is a form of t-test is formula is given by

C.D = t.S.E(d)

Where SE = Standard Error

$$S.E(d) = \sqrt{EMS\left(\frac{1}{r_i} - \frac{1}{r_j}\right)}$$

EMS = Error Mean Square

In the case of same replication the standard is $S.E = \sqrt{\frac{2EMS}{r}}$

In this formula t is the critical (table) value of t for a specified level of significance and error degrees of freedom $r_i$ and $r_j$ for the number of replications for the $i^{th}$ and $j^{th}$ treatment respectively, the formula for t-test is

$$t = \frac{Y_i - Y_j}{S\sqrt{\frac{1}{r_i} + \frac{1}{r_j}}}$$

The two treatment means are declared significantly different at specified level of significance.

If the difference exceeds the calculated CD value, otherwise they are not significant CD value.

### 3.2.2 Duncan's multiple range test (DMRT)

In a set of t-treatments if the comparison of all possible pairs of treatment mean is required. We can use Duncan's multiple range test. The DMRT can be used irrespective of whether F is significant or not.

Procedure:

Step: 1

Arrange the treatments in descending order that is to range.

Step: 2

Calculate the S.E of mean as

$$S.E(\bar{Y}) = \sqrt{\frac{SQ^2}{r}} = \sqrt{\frac{EMS}{r}}$$

Step: 3

From statistical table write the significant student zed range as (rp), p = 1,2,.....t treatment and error degrees of freedom.

Step: 4

Calculate the shortest significance range as $R_p$ where $R_p = r_p.S.E(\bar{Y})$

Step: 5

From the largest mean subtract the Rp for largest P. Declare as significantly different from the largest mean. For the remaining treatment whose values are larger than the difference (largest mean-largest Rp). Compare the difference with appropriate Rp value.

Step: 6

Continue this process till all the treatment above.

Step: 7

Present the results by using either the line notation (or) the alphabet notation to indicate which treatment pair which are significantly different from each other.

**3.2.3 Tukey's range test:**

Tukey's range test is also known as Tukey's test, Tukey's HSD (Honest significance difference) test. It can be used on raw data or in cons unction with an ANOVA (post-hoc analysis) to find means that are significantly different from each other. Tukey's test compares the means of every treatment to the means of every other treatment.

The test statistic:

Tukey's test is based on a formula very similar to that of the t-test. In fact, Tukey's test is essentially a t-test, except that is corrects for experiment wise error rate.

Formula to,

$$q_s = \frac{Y_A - Y_B}{S.E}$$

Where $Y_A$ is a larger of the two means being compared. $Y_B$ is the smaller of the two means being compared. S.E is the standard error.

This qs value can then be compared to a q value from the studentized range distribution. If the $q_s$ value is larger than the q critical value obtained from the distribution. The two means are said to significantly different.

The studentized range distribution:

$$q = \frac{(\bar{y}_{\max} - \bar{y}_{\min})}{S/\sqrt{2/n}}$$

**3.2.4 Student – Newman Keuls (SNK) test**

The Newman-Keuls (or) student Newman Keuls (SNK) method is a stepwise multiple comparison. Procedure used to identify sample means that are significantly different from each other. It was named after student (1927) D. Newman and M. Keuls,

Procedure:

1. The Newman Keuls method employs stepwise approach when comparing sample means.
2. Prior to any mean comparison, all sample means are rank ordered in ascending or descending order there by producing an ordered range (p) of sample means.
3. A comparison is then made between the largest and smallest sample means within the largest range.
4. Assuming that the largest range is four means (p=4) a significant difference between the largest and smallest means as revealed by the Newman-Keuls

method would result in a reflection of the null hypothesis for that specific range of means.

5. The next largest comparison of two sample means would then be made within a smaller range of three means (p=3).
6. Continue this process until a final comparison is made.
7. It there is no significant difference between the two sample means. Then all the null hypothesis within that range would be retained and no further comparisons within smaller ranges are necessary.

### 3.3 Analysis of Covariance for two way classification (Random Block Design) with one concomitant variable

Suppose we want to compare v treatments, each treatment replicated r times so that total number of experimental units is n = vr. Suppose that the experiment is conducted with Randomized Block Design(RBD) layout.

Assuming a linear relationship between the response variable (y) and concomitant variable(x) the appropriate statistical model for ANOCOVA for RBD(with one concomitant variable) is:

$$y_{ij} = \mu + \alpha_i + \theta_j + \beta\left(x_{ij} - \bar{x}_{..}\right) + e_{ij} \qquad \ldots(3.1)$$

Where μ is the general mean effect

$\alpha_i$ is the (fixed) additional effect due to the $i^{th}$ treatment ,(i=1,2,…,v)

$\theta_j$ is the (fixed) additional effect due to the $j^{th}$ block ,(j=1,2,…,r)

β is the coefficient of regression of y on x

$x_{ij}$ is the value of the concomitant variable corresponding to the response variable $y_{ij}$ and $e_{ij}$ is the random error effect so that

$$\sum_{i=1}^{v} \alpha_i = 0,\ \sum_{j=1}^{r} \theta_j = 0\ e_{ij} \overset{i.i.d}{\sim} N(0, \sigma_e^2)$$

Estimation of parameters in (1) we shall estimate the parameters μ, $\alpha_i$ ((i=1,2,…,v), $\theta_j$ (j=1,2,…,r) and β, using the principle of least squares by minimizing the error sum of squares in (1)

$$SSE = \sum_i \sum_j e_{ij}^2 = \sum_{i=1}^{v} \sum_{j=1}^{r} \left[ y_{ij} - \mu - \alpha_i - \theta_j - \beta\left(x_{ij} - \bar{x}_{..}\right) \right]^2$$

...(3.2)

Normal equations for estimating the parameters are

$$\frac{\partial (SSE)}{\partial \mu} = 0 = -2 \sum_i \sum_j \left[ y_{ij} - \mu - \alpha_i - \theta_j - \beta (x_{ij} - \bar{x}_{..}) \right]$$

$$\sum_i \sum_j y_{ij} - \sum_i \sum_j \mu - \sum_i \sum_j \alpha_i - \sum_i \sum_j \theta_j - \sum_i \sum_j \beta\left(x_{ij} - \bar{x}_{..}\right) = 0$$

$$\sum_i \sum_j y_{ij} - rv\mu - r \sum_i \alpha_i - v \sum_j \theta_j - \sum_i \sum_j \beta\left(x_{ij} - \bar{x}_{..}\right) = 0$$

...(3.3)

$$\frac{\partial (SSE)}{\partial \alpha_i} = 0 = -2 \sum_j (y_{ij} - \mu - \alpha_i - \theta_j - \beta (x_{ij} - \bar{x}..)$$

$$\sum_j y_{ij} - \sum_j \mu - \sum_j \alpha_i - \sum_j \theta_j - \sum_j \beta (x_{ij} - \bar{x}..) = 0$$

$$\sum_j y_{ij} - \sum_j \mu - r\alpha_i - \sum_j \theta_j - \sum_j \beta (x_{ij} - \bar{x}..) = 0 \quad ...(3.4)$$

$$\frac{\partial (SSE)}{\partial \theta_j} = 0 = -2 \sum_i (y_{ij} - \mu - \alpha_i - \theta_j - \beta (x_{ij} - \bar{x}..)$$

$$\sum_i y_{ij} - \sum_i \mu - \sum_i \alpha_i - \sum_i \theta_j - \sum_i \beta (x_{ij} - \bar{x}..) = 0$$

$$\sum_i y_{ij} - \sum_i \mu - \sum_i \alpha_i - v\theta_j - \sum_i \beta (x_{ij} - \bar{x}..) = 0 \qquad ...(3.5)$$

$$\frac{\partial (SSE)}{\partial \beta} = -2 \sum_i \sum_j [\{ y_{ij} - \mu - \alpha_i - \theta_j - \beta (x_{ij} - \bar{x}_{..}) \} (x_{ij} - \bar{x}_{..})] ...(3.6)$$

From equation (3. 2)

$$\hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{rv} = \bar{y}_{..}$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}..) = 0$$

From equation (3. 3)

$$\sum_i y_{ij} - r(\hat{\mu} + \hat{\alpha}_i) - \hat{\beta} \sum_j (x_{ij} - \bar{x}_{..}) = 0$$

$$r\bar{y}_{i.} - r(\bar{y}_{..} + \hat{\alpha}_i) - \hat{\beta}.r(\bar{x}_{i.} - \bar{x}_{..}) = 0$$

$$\bar{y}_{i.} - (\bar{y}_{..} + \hat{\alpha}_i) - \hat{\beta}.(\bar{x}_{i.} - \bar{x}_{..}) = \frac{0}{r} = 0$$

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \hat{\beta}(\bar{x}_{i.} - \bar{x}..)$$

From equation (3.4)

$$\sum_i y_{ij} - v\mu - v\theta_j - v\hat{\beta}(x_{.j} - \bar{x}_{..}) = 0$$

$$v\bar{y}_{.j} - v(\mu + \theta_j) - v\hat{\beta}(x_{.j} - \bar{x}_{..}) = 0$$

$$\bar{y}_{.j} - (\mu + \theta_j) - \hat{\beta}(x_{.j} - \bar{x}_{..}) = \frac{0}{v} = 0$$

$$\bar{y}_{.j} - \mu - \theta_j) - \hat{\beta}(x_{.j} - \bar{x}_{..}) = 0$$

$$\hat{\theta}_j = (\bar{y}_{.j} - \bar{y}_{..}) - \hat{\beta}(x_{.j} - \bar{x}_{..})$$

Substituting these estimated values in equation (3. 5)

$$0=\sum_i\sum_j[\{y_{ij}-\bar{y}_{..}\{\bar{y}_{i.}-\bar{y}_{..}-\hat{\beta}(\bar{x}_{i.}-\bar{x}_{..})\}-\{\bar{y}_{.j}-\bar{y}_{..}-\hat{\beta}(\bar{x}_{.j}-\bar{x}_{..})\}-\hat{\beta}(x_{ij}-\bar{x})\}(x_{ij}-\bar{x}_{..})]$$

$$0=\sum_i\sum_j[\{(y_{ij}-\bar{y}_{i.}-\bar{y}.j+\bar{y}_{..})-\hat{\beta}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})\}(x_{ij}-\bar{x}_{..})$$

$$\sum_i\sum_j[\{(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})-\hat{\beta}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})\}$$
$$\times\{(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})+(\bar{x}_{i.}-\bar{x}_{..})+(\bar{x}_{.j}-\bar{x}_{..})\}]=0$$

$$\sum_i\sum_j(y_{ij}-\bar{y}_{i.}\cdot-\bar{y}_{.j}+\bar{y}_{..})(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}-\bar{x}_{..})-\hat{\beta}\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2=0$$

(the product terms will be zero since algebraic sum of deviations from mean is zero)

$$\hat{\beta}=\frac{\sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})}{\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2}$$

Let us write:

$$E_{xx}=\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})^2 \;\; ; \;\; E_{yy}=\sum_i\sum_j(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y})^2 \quad \text{and}$$

$$E_{xy}=\sum_i\sum_j(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x}_{..})(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})$$

Then, $\hat{\beta}=\frac{E_{xy}}{E_{xx}}$

Substituting the values of $\hat{\mu},\hat{\alpha}_i,\hat{\theta}_j$ *and* $\hat{\beta}$ in (3.2), the unrestricted error sum of squares for model (3.1) becomes:

SSE = Minimum value of error S.S

$$SSE=\sum_i\sum_j e_{ij}^2=\sum_{i=1}^{v}\sum_{j=1}^{r}\left[y_{ij}-\hat{\mu}-\hat{\alpha}_i-\hat{\theta}_j-\hat{\beta}\left(x_{ij}-\bar{x}_{..}\right)\right]^2$$

$$= \sum_i \sum_j [(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) - \hat{\beta}(x_{ij} - \bar{x}_i - \bar{x}_{.j} + \bar{x}_{..})]^2 \quad \text{(on simplification)}$$

$$= \sum_i \sum_j [(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 + \hat{\beta}^2 \sum_i \sum_j (x_{ij} - \bar{x}_i - \bar{x}_{.j} + \bar{x}_{..})^2 - 2\hat{\beta} \sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

$$= E_{yy} + \left(\frac{E_{xy}}{E_{xx}}\right)^2 .E_{xx} - 2\frac{E_{xy}}{E_{xx}} .E_{xy}$$

$$= E_{yy} - \frac{E_{xy}^2}{E_{xx}} = E_{yy} - \hat{\beta} E_{xy}$$

$$= Error\ S.S\ for\ y\ in\ RBD - \frac{E_{xy}^2}{E_{xx}}$$

Since $E_{xy}^2/E_{xx}>0$, there is reduction in SSE if we apply ANOCOVA to RBD, the reduction

d.f for SSE = Total d.f.-(d.f. due to treatments) – (d.f. due to blocks) – (d.f. due to β )

=(rv-1)-(v-1)-(r-1)-1 = (r-1)(v-1)-1

Under the null hypothesis:

$H_0$: All treatment effects are equal, $H_0$: $\alpha_1=\alpha_2=\ldots=\alpha_v=0$,

The model reduces to

$$y_{ij} = \mu + \alpha_i + \theta_j + \hat{\beta}(x_{ij} - \bar{x}_{..}) + e_{ij}$$

Restricted error sum of squares under $H_0$ is given by

$$(SSE)^* = \sum_i \sum_j e_{ij}^2 = \sum_{i=1}^{v} \sum_{j=1}^{r} [y_{ij} - \mu - \alpha_i - \theta_j - \beta'(x_{ij} - \bar{x}_{..})]^2 \qquad \ldots (3.7)$$

The normal equations for estimating μ, $\theta_j$ and β' are given by:

$$\frac{\partial (SSE)^*}{\partial \mu} = 0 = -2 \sum_i \sum_j [y_{ij} - \mu - \alpha_i - \theta_j - \beta'(x_{ij} - \bar{x}_{..})]$$

$$\sum_i \sum_j y_{ij} - \sum_i \sum_j \mu - \sum_i \sum_j \alpha_i - \sum_i \sum_j \theta_j - \sum_i \sum_j \beta'(x_{ij} - \bar{x}_{..}) = 0$$

$$\sum_i \sum_j y_{ij} - rv\mu - r\sum_i \alpha_i - v\sum_j \theta_j - \sum_i \sum_j \beta'\left(x_{ij} - \bar{x}_{..}\right) = 0 \quad \ldots(3.8)$$

$$\frac{\partial (SSE)^*}{\partial \alpha_i} = 0 = -2\sum_j (y_{ij} - \mu - \alpha_i - \theta_j - \beta'(x_{ij} - \bar{x}..)$$

$$\sum_j y_{ij} - \sum_j \mu - \sum_j \alpha_i - \sum_j \theta_j - \sum_j \beta'(x_{ij} - \bar{x}..) = 0$$

$$\sum_j y_{ij} - \sum_j \mu - r\alpha_i - \sum_j \theta_j - \sum_j \beta'(x_{ij} - \bar{x}..) = 0 \quad \ldots(3.9)$$

$$\frac{\partial (SSE)}{\partial \theta_j} = 0 = -2\sum_i (y_{ij} - \mu - \alpha_i - \theta_j - \beta(x_{ij} - \bar{x}..)$$

$$\sum_i y_{ij} - \sum_i \mu - \sum_i \alpha_i - \sum_i \theta_j - \sum_i \beta'(x_{ij} - \bar{x}..) = 0$$

$$\sum_i y_{ij} - \sum_i \mu - \sum_i \alpha_i - v\theta_j - \sum_i \beta'(x_{ij} - \bar{x}..) = 0 \quad \ldots(3.10)$$

$$\frac{\partial (SSE)^*}{\partial \beta'} = -2\sum_i \sum_j [\{y_{ij} - \mu - \alpha_i - \theta_j - \beta'(x_{ij} - \bar{x}_{..})\}(x_{ij} - \bar{x}_{..})] \quad \ldots(3.11)$$

From equation (3.8)

$$\hat{\mu} = \frac{\sum_i \sum_j y_{ij}}{rv} = \bar{y}_{..}$$

$$= \sum_i \sum_j (x_{ij} - \bar{x}..) = 0$$

From equation (3.9)

$$\sum_i y_{ij} - r(\hat{\mu} + \hat{\alpha}_i) - \hat{\beta}' \sum_j (x_{ij} - \bar{x}_{..}) = 0$$

$$r\bar{y}_{i.} - r(\bar{y}_{..} + \hat{\alpha}_i) - \hat{\beta}' r(\bar{x}_{i.} - \bar{x}_{..}) = 0$$

$$\bar{y}_{i.} - (\bar{y}_{..} + \hat{\alpha}_i) - \hat{\beta}'(\bar{x}_{i.} - \bar{x}_{..}) = \frac{0}{r} = 0$$

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \hat{\beta}'(\bar{x}_{i.} - \bar{x}..)$$

From equation (3.10)

$$\sum_i y_{ij} - v\mu - v\theta_j - v\hat{\beta}(x_{.j} - \bar{x}_{..}) = 0$$

$$v\bar{y}_{.j} - v(\mu + \theta_j) - v\hat{\beta}(x_{.j} - \bar{x}_{..}) = 0$$

$$\bar{y}_{.j} - (\mu + \theta_j) - \hat{\beta}'(x_{.j} - \bar{x}_{..}) = \frac{0}{v} = 0$$

$$\bar{y}_{.j} - \mu - \theta_j - \hat{\beta}'(x_{.j} - \bar{x}_{..}) = 0$$

$$\hat{\theta}_j = (\bar{y}_{.j} - \bar{y}_{..}) - \hat{\beta}'(x_{.j} - \bar{x}_{..})$$

From equation (3.11), we get

$$\sum_i \sum_j [(x_{ij} - \bar{x}_{..})\{y_{ij} - \bar{y}_{..} - (\bar{y}_{.j} - \bar{y}_{..}) + \hat{\beta}'(\bar{x}_{.j} - \bar{x}_{..}) - \hat{\beta}'(x_{ij} - \bar{x}_{..})\}] = 0$$

$$\sum_i \sum_j [x_{ij} - \bar{x}_{..})\{y_{ij} - \bar{y}_{.j}) - \hat{\beta}$$

$$\sum_i \sum_j [(x_{ij} - \bar{x}_{.j} + \bar{x}_{.j} - \bar{x}_{..})\{(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'(x_{ij} - \bar{x}_{.j})\}] = 0$$

$$\sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'(x_{ij} - \bar{x}_{.j})^2 = 0, \textit{the other product terms are zero}$$

$$\hat{\beta}' = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j})}{\sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2}$$

*Let us define*:

$$E_{xx}' = \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2; E_{yy}' = \sum_i \sum_j (y_{ij} - \bar{y}_{.j})^2; E_{xy}' = \sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j})$$

$$\hat{\beta}' = \frac{E_{xy}'}{E_{xx}'}$$

Hence, under $H_0$, the restricted error sum of squares is given by:

$(SSE)^*$ =minimum value of error S. S

$$= \sum_i \sum_j \left[y_{ij} - \hat{\mu} - \hat{\theta}_j - \hat{\beta}'(x_{ij} - \bar{x}_{..})\right]^2$$

$$\sum_i \sum_j \left[(y_{ij} - \bar{y}_{.j}) - \hat{\beta}'(x_{ij} - \bar{x}_{.j})\right]^2$$

$$\sum_i \sum_j (y_{ij} - \bar{y}_{.j})^2 + \hat{\beta}'^2 \sum_i \sum_j (x_{ij} - \bar{x}_{.j})^2 - 2\hat{\beta}' \sum_i \sum_j (x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{..})$$

$$= E_{yy}' + \left(\frac{E_{xy}'}{E_{xx}'}\right)^2 E_{xx}' - 2\left(\frac{E_{xy}'}{E_{xx}'}\right)(E_{xy}')$$

$$= E_{yy}' - \frac{E_{xy}'^2}{E_{xx}'} \qquad \ldots 3.12$$

d.f. for$(SSE)^*$ =total d.f. –d.f. for Blocks –d.f. for β

=(vr-1)-(r-1)-1=vr-r-1=r(v-1)-1

Adjusted sum of squares for treatments (SST = $S_t^2$) is given by

$(SST = S_t^2)=(SSE)^* -(SSE)$

Where (SSE) and $(SSE)^*$ are given in (3.1) and (6.12) respectively.

d.f.(SST) =d.f.$(SSE)^*$ - d.f.(SSE)

=[r(v-1)-1]-[(r-1)(v-1)-1]=v-1

$$MST(s_t^2) = \frac{SST}{d.f.} = \frac{S_t^2}{d.f.} = \frac{(SSE)^* - (SSE)}{v-1}$$

Hence the test statistic for testing $H_0$: $\alpha_1=\alpha_2= \ldots=\alpha_v$, is given by

$$F = \frac{MST}{MSE} = \left[\frac{(SSE)^* - SSE}{(v-1)}\right] \times \frac{(r-1)(v-1)-1}{SSE} \sim F_{v-1,(r-1)(v-1)-1}$$

If $F > F_{v-1,\,(r-1)(v-1)-}(\alpha)$ then $H_0$ is rejected at $\alpha$ level of significance, otherwise we fail to reject $H_0$.

ANOCOVA Table for RBD. Let us write:

$SS_{yy} = E_{yy}+T_{yy}+B_{yy}$

Where $T_{yy} = r\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2$ is the treatment S.S for y for RBD

$B_{yy} = v\sum_i(\bar{y}_{.j} - \bar{y}_{..})^2$ is the block S.S for y for RBD

$E_{yy} = \sum_i\sum_j(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ is the error S.S for y for RBD

Similarly we have,

$SS_{xx}=E_{xx}+T_{xx}+B_{xx}$

$SP_{xy}=E_{xy}+T_{xy}+B_{xy}$

Where $T_{xy} = \sum_i r(\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})$ *and* $B_{xy} = \sum_j v(\bar{x}_{.j} - \bar{x}_{..})(\bar{y}_{.j} - \bar{y}_{..})$

$E_{xx}= \sum_i\sum_j(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$ ; $E_{yy} = \sum_i\sum_j(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y})^2$ and

$$E_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

Using the above notations, the above statistical analysis can be elegantly expressed in the ANOCOVA table 3.1:

**Table 3.1: ANOCOVA Table (RBD)**

| Sources of variation | d.f. | Sum of Squares and Products | | | Estimate of β | Adjusted $SS_{yy}$ | Adjusted d.f. |
|---|---|---|---|---|---|---|---|
| | | $SS_{xx}$ | $SS_{yy}$ | $SP_{xy}$ | | | |
| Blocks | r-1 | $B_{xx}$ | $B_{yy}$ | $B_{xy}$ | | | |
| Treatments | v-1 | $T_{xx}$ | $T_{yy}$ | $T_{xy}$ | | | |
| Error | (r-1)(v-1)-1 | $E_{xx}$ | $E_{yy}$ | $E_{xy}$ | $\hat{\beta} = \frac{E_{xy}}{E_{xx}}$ | SSE | (r-1)(v-1)-1 |
| Treatment+ Error | r(v-1) | $E_{xx}'$ | $E_{yy}'$ | $E_{xy}'$ | $\hat{\beta}' = \frac{E_{xy}'}{E_{xx}'}$ | $SSE^*$ | r(v-1)-1 |
| | | | | | | SSE-$SSE^*$ | v-1 |

Note that $E_{xx}'=T_{XX}+E_{xx}$; $E_{yy}'=T_{yy}+E_{yy}$ and $E_{yy}'=T+Exy$

**Unit -IV**

## 4.1. Factorial Experiments

So far we have discussed the experiments in which the effects of a single set of treatments were estimated and compared. In these experiments most of the other variable factors were kept constant. In practice, the response of biological organisms to the factor of interest is expected to differ under different levels of other factors. For example, the yield of paddy varieties may differ under different rates of fertilizer application, spacing and irrigation schedules. Hence, in agricultural research we frequently wish to what happens with a range of combination of factors. When several factors are investigated simultaneously in a single experiment, such experiments are known as factorial experiments.

**Advantages of Factorial Experiment**

1. It increases the scope of the experiment and its inductive value and it does so mainly by giving information not only on the main factors but on their interactions.
2. The various levels of one factor constitute replications of other factors and increase the amount of information obtained on all factors.
3. When there are no interactions, the factorial design gives the maximum efficiency in the estimate of the effects.
4. When interactions exist, their nature being unknown a factorial design is necessary to avoid misleading conclusions.
5. In the factorial design the effect of a factor is estimated at several levels of other factors and the conclusions hold over a wide range of conditions.

   Basic Ideas and Notations in the $2^n$ Factorial Experiment

   Let us first consider the design of the of the form $2^n$ in which there are n factors, each at two levels. Levels may be quite qualitative alternatives like two species of a plant. In some cases one level is simply the control group, ie., the absence of the factor and the other is its presence.

   In order to develop extended notation to present the analysis of the design in a concise form, let us start, for simplicity with a $2^2$ – factorial design.

## 4.2. $2^2$ – Factorial Design

Here we have two factors each at two levels (0,1), say, so that there are 2×2=4 treatment combinations in all. Following the notations due to Yates, let the capital letters A and B indicates the names of the two factors under study and let the small letters a and b denote one of the

two levels of each of the corresponding factors and this will be called the second level. The first level of A and B is generally expressed by the absence of the corresponding letter in the treatment combinations. The four treatment combinations can be enumerated as follows:

$a_0b_0$ or 1 : Factors A and B, both at first level.
$a_1b_0$ or a : A at second level and B at first level
$a_0b_1$ or b : A at first level and B at second level
$a_1b_1$ or ab : A and B both at second level

These four treatment combinations can be compared by laying out the experiment (i) R. B.D ., with r replicates (say), each replicate containing 4 units or (ii) 4 × 4 L. S.D., and ANOVA can be carried out accordingly. In the above cases there are 3 d.f associated with the treatment effects. In factorial experiment our main objective is to carry out separate tests for the main effects A, B and the interaction AB, splitting the treatment S.S with 3 d.f into three orthogonal components each with 1 d.f and each associated either with the main effects A and B or interactions AB.

### 4.2.1Main and interaction effects of $2^2$ factorial design:

Suppose the factorial experiment $\mathbf{2^2}$=4 Treatment is conducted 'r' Blocks (or)Replicator.

Let [1]: total yield of the r units receiving the treatment 1.

Let [a]: Total yield of the r units receiving the treatment a.

Let [b]: Total yield of the r units receiving the treatment b.

Let [ab]: Total yield of the r units receiving the treatment ab.

(1): $\frac{[1]}{r}$ = the mean yields of the r units receiving the treatment 1.

(a): $\frac{[a]}{r}$ = the mean yields of the r units receiving the treatment a.

(b): $\frac{[b]}{r}$ = the mean yields of the r units receiving the treatment b.

(ab): $\frac{[ab]}{r}$ = the mean yields of the r units receiving the treatment ab.

The effect of A at the first level $b_0$ of B = $(a_1b_0)-(a_0b_0)$

= (a)-(1) --------> 1

The effect of A at the first level $b_1$ of B = $(a_1b_1)-(a_0b_1))$

= (ab)-(b) -------> 2

The effect of B at the first level $a_0$ of A = $(a_0b_1)-(a_0b_0)$

= (b)-(1) ------> 3

The effect of B at the first level $a_1$ of A = $(a_1b_1)-(a_1b_0)$

= (ab)-(a) ----- > 4

The main effect due to A is defined by

(2)+ (1) => A= $\frac{1}{2}$ [(ab)-(b) + (a)-(1)]

= $\frac{1}{2}$ [b (a-1) +1(a-1)]

= $\frac{1}{2}$ [(a-1) + (b+1)]

The main effect due to B is defined by

(4)+ (3) => B= $\frac{1}{2}$ [(b)-(1) + (ab)-(a)]

= $\frac{1}{2}$ [(b-1) +a (b-1)]

= $\frac{1}{2}$ [(a+1) (b-1)]

Interaction effect due to AB is defined by

(2)- (1) => AB= $\frac{1}{2}$ [(ab)-(b)-{(a)-(1)}]

= $\frac{1}{2}$ [b (a-1)-1(a-1)]

$$= \frac{1}{2}[(a-1)(b-1)] \text{ ---------- > 5}$$

Interaction effect due to BA is defined by

(4)- (3) => $BA = \frac{1}{2}[(ab)-(a)-\{(b)-(1)\}]$

$$= \frac{1}{2}[a(b-1)-1(b-1)]$$

$$= \frac{1}{2}[(a-1)(b-1)] \quad \text{---------- > 6}$$

The equation (5) and (6) are same hence the interaction effect AB as same as BA.

### 4.2.3. Statistical Analysis for $2^2$ Factorial experiment

Factorial Experiment are conducted either CRD, RBD, and LSD thus they can be Analysis in the usual manner except that is case treatment sum of sequence split into three orthogonal components each i.d.f. It has been already be pointed out the main effects A and B and the interaction AB. The sum of squares due to the Factorial effects A, B, and AB is obtained by multiplying by the squares of the practice these effects are usually computed from the treatment total [a], [b] and [ab] etc. The Factorial effects totals are given by the expression.

[A] = [ab]-[b] + [a]-[1]

[B] = [ab]-[a] + [b]-[1]

[AB] = [ab]-[b]- [a] + [1]

The sum of squares due to any Factorial effect is obtained by multiplying the square of the effect total by the Factor $\frac{1}{4r}$. Where r is the common replicate number sum of square due to main effect of A= $\frac{[A]^2}{4r} = S_A^2$

Similarly S.S due to main effects of B = $\frac{[B]^2}{4r} = S_B^2$

S.S due to interaction effect of AB= $\frac{[AB]^2}{4r} = S_{AB}^2$

Each with 1.D.F

Correction factor (C.F) = $\frac{G^2}{N}$

Total sum of square = $\sum_{ij \in s} y_{ij}^{\ 2} - C.F = S_T^2$

Block S.S = $\frac{\sum B_j^2}{4} - C.F = S_R^2$

S.S Treatment = $\frac{\sum T_i^2}{r} - C.F = S_t^2$

S.S Treatment = Total S.S – [S.SB+ S.ST] = $S_E^2$

**Table 4.1: ANOVA Table**

| **Source of variation** | **Sum of square** | **Degrees of freedom** | **Mean sum of square** | **F- ratio** |
|---|---|---|---|---|
| Block | $S_R^2$ | r-1 | $\frac{S_R^2}{r-1} = S_R^2$ | $\frac{S_R^2}{S_E^2} = F_R$ |
| Treatment | $S_t^2$ | 3 | $\frac{S_t^2}{3} = S_t^2$ | $\frac{S_t^2}{S_E^2} = F_t$ |
| Main Factor A | $S_A^2$ | 1 | $\frac{S_A^2}{1} = S_A^2$ | $\frac{S_A^2}{S_E^2} = F_A$ |
| Main Factor B | $S_B^2$ | 1 | $\frac{S_B^2}{1} = S_B^2$ | $\frac{S_B^2}{S_E^2} = F_B$ |

| | | | | |
|---|---|---|---|---|
| Interaction effect AB | $S_{AB}^2$ | 1 | $\frac{S_{AB}^2}{1} = S_{AB}^2$ | $\frac{S_{AB}^2}{S_E^2} = F_{AB}$ |
| Error | $S_E^2$ | 3(r-1)<br>$(2^2\text{-}1)(\text{r-}1)$ | $\frac{S_E^2}{3(r-1)} = S_E^2$ | |
| Total | $S_T^2$ | 4r-1<br>$(2^2\,\text{r-}1)$ | | |

**Table value:**

F table [(r-1), 3(r-1)] d.f at 5% level = $F_R^*$

F table [3, 3(r-1)] d.f at 5% level = $F_t^*$

F table [1, 3(r-1)] d.f at 5% level = $F^*$

**Conclusion:**

(i) $F_R < F_R^*$ we need not reject and we conclude that there is no significance difference between replicates.

(ii) $F_t < F_t^*$ we need not reject and we conclude that there is no significance difference between Treatment.

(iii) $F_A < F_A^*$ we need not reject and we conclude that there is no significance

(iv) $F_B < F_B^*$ we need not reject and we conclude that there is no significance difference main effect B.

(v) $F_{AB} < F^*$ we need not reject and we conclude that there is no significance difference interaction effect AB.

**Table 4.2: Yates Method for $2 \times 2$ Factorial experiment:**

| **Treatment combination** | **Total yield from all replicates (2)** | **(3)** | **(4)** | **Effect total** |
|---|---|---|---|---|
| 1 | [1] | [1]+[a] | [1]+[a]+[b]+[ab] | Ground total |
| A | [a] | [b]+[ab] | [a]-[1]+[ab]-[b] | [A] |
| B | [b] | [a]-[1] | [b]+[ab]-[1]-[a] | [B] |
| Ab | [ab] | [ab]-[b] | [ab]-[b]+[a]-[1] | [AB] |

**Problem 1**:

Find out ,the effect and interaction effect in the following 2 Factorial experiment and write dawn the analysis of variance table;-

**Table 4.3 ANOVA Table**

| Block | Treatment | (a) | (b) | (ab) |
|---|---|---|---|---|
| | 00 | 10 | 01 | 11 |
| I | 64 | 25 | 30 | 10 |
| II | 25 | 14 | 50 | 33 |

| | | | | |
|---|---|---|---|---|
| III | 76 | 12 | 41 | 17 |
| IV | 75 | 33 | 25 | 10 |

**Solution:**

**Null hypothesis $H_{ob}$:**

There is no significance difference between blocks.

$H_{ot}$: There is no significance difference between Treatments.

$H_{oA}$: There is no significance difference between main effect A.

$H_{oB}$: There is no significance difference between main effect B.

$H_{oAB}$: There is no significance difference between effect AB.

| Treatment Combination | Total yield from all blocks | (3) | Effects Total | S.S |
|---|---|---|---|---|
| 1 | 240 | 324 | 540 | $\frac{(540)^2}{16} = 18225$ |
| a | 84 | 216 | -232 | $\frac{(-232)^2}{16} = 3364$ |
| b | 146 | -156 | -108 | $\frac{(-108)^2}{16} = 729$ |
| ab | 70 | -76 | 80 | $\frac{(80)^2}{16} = 400$ |

C.F = $\frac{G^2}{N} = \frac{(540)^2}{16} = 18225$

C.F = 18225

Total sum of square = $\sum_{ij \in s} y_{ij} - C..F = (64)^2 + (25)^2 + .....(25)^2 + (10)^2 - C..F$

= 25460- 18225

= 7235

Sum of square due to Block = $\frac{\sum B_j^2}{4} - C.F$

= $\frac{73290}{4} - 18225$

S.SB = 97.5

Sum of square due to Treatment = $\frac{\sum T_i^2}{4} - C.F = \frac{90872}{4} - 18225$

S.S.$T_r$ = 4493

Sum of square due to error = T.S.S – (S.S.B + S.S.$T_r$)

= 7235- (97.5+4493)

S.S.E = 2633.5

**Table 4.3: Yates method for $2^2$ Factorial experiment:**

| Source of variation | Sum of square | Degree of freedom | Mean sum of square | F- ratio |
|---|---|---|---|---|
| Block | 97.5 | 3 | 32.5 | 0.1106 $F_R$ |
| Treatments | 4493 | 3 | 1497.67 | 5.0975 $F_T$ |
| Main effect A | 3364 | 1 | 3364 | 11.4499 $F_A$ |

| | | | | |
|---|---|---|---|---|
| Main effect B | 729 | 1 | 729 | 2.48 $F_B$ |
| Interaction effect of AB | 400 | 1 | 400 | 1.3614 $F_{AB}$ |
| Error | 2644.5 | 9 | 293.8 | - |
| Total | 7235 | 15 | - | - |

**F- Table value:**

F (3.9) d.f at 5% level = 3.86 → $F_A^*$

F (1.9) d.f at 5% level = 5.12→ $F_B^*$

**Conclusion:**

(i) $F_R < F_A^*$ we need not reject $H_0$ and we conclude that there is no significance difference between blocks.

(ii) $F_T > F_A^*$ we need not accept $H_0$ and we conclude that there is no significance difference between Treatments.

(iii) $F_A > F_B^*$ We need not accept $H_0$ and we conclude that there is no significance difference between main effect A.

(iv) $F_B < F_B^*$ We need not reject $H_0$ and we conclude that there is no significance difference between main effect B.

(v) $F_{AB} < F_B^*$ We need not reject $H_0$ and we conclude that there is no significance difference between interaction effect AB.

### 4.3. $2^3$ Factorial Experiment

( main and interaction effect of $2^3$ factorial experiment) suppose that are factorial experiment $2^3$=8 Treatment combinations is conducted 'r' Blocks (or) replicates.

[1] Total of r units receiving the treatment 1.

[a] Total of r units receiving the treatment a.

[b] Total of r units receiving the treatment b.

[ab] Total of r units receiving the treatment ab.

[c] Total of r units receiving the treatment c.

[ac] Total of r units receiving the treatment ac.

[bc] Total of r units receiving the treatment bc.

[abc] Total of r units receiving the treatment abc.

**Simple effect of A:-**

| Level of B | Level of C | Effect of A |
|---|---|---|
| $b_0$ | $c_0$ | $a_1b_0c_0 - a_0b_0c_0$ = (a)-(1) |
| $b_1$ | $c_0$ | $a_1b_1c_0 - a_0b_1c_0$ = (ab)-(b) |

| | | |
|---|---|---|
| $b_0$ | $c_1$ | $a_1b_0c_1$ - $a_0b_0c_1$ = (ac)-(c) |
| $b_1$ | $c_1$ | $a_1b_1c_1$ - $a_0b_1c_1$ = (abc)-(bc) |

**Simple effect of A**

$$= \frac{1}{4}[(abc)-(bc)+(ac)-(c)+(ab)-(b)+(a)-(1)]$$

$$= \frac{1}{4}[bc(a-1)+c(a-1)+b(a-1)+1(a-1)]$$

$$= \frac{1}{4}[(a-1)(bc+c+b+1)]$$

$$= \frac{1}{4}[(a-1)\{c(b+1)+1(b+1)\}]$$

$$= \frac{1}{4}[(a-1)\{(b+1)(c+1)\}]$$

$$= \frac{1}{4}[(a-1)(b+1)(c+1)]$$

S.E. of A = $\frac{1}{4}[(a-1)(b+1)(c+1)]$

**Simple effect of B:**

| Level of A | Level of C | Effect of B |
|---|---|---|
| $a_0$ | $c_0$ | $a_0b_1c_0$ - $a_0b_0c_0$ = (b)-(1) |
| $a_1$ | $c_0$ | $a_1b_1c_0$ - $a_1b_0c_0$ = (ab)-(a) |
| $a_0$ | $c_1$ | $a_0b_1c_1$ - $a_0b_0c_1$ = (ac)-(a) |

$a_1$

$c_1$ $a_1b_1c_1$ - $a_1b_0c_1$ = (abc)-(ac)

**Simple effect of B**

$$= \frac{1}{4}[(abc)-(ac)+(bc)-(c)+(ab)-(a)+(b)-(1)]$$

$$= \frac{1}{4}[ac(b-1)+c(b-1)+a(b-1)+1(b-1)]$$

$$= \frac{1}{4}[(b-1)(ac+c+a+1)]$$

$$= \frac{1}{4}[(b-1)\{c(a+1)+1(a+1)\}]$$

$$= \frac{1}{4}[(b-1)\{(a+1)(c+1)\}]$$

$$= \frac{1}{4}[(b-1)(a+1)(c+1)]$$

S.E. of B $= \frac{1}{4}[(b-1)(a+1)(c+1)]$

**Simple effect of C:-**

| Level of A | Level of B | Effect of C |
|---|---|---|

| | | |
|---|---|---|
| $a_0$ | $b_0$ | $a_0b_0c_1 - a_0b_0c_0$ = (c)-(1) |
| $a_1$ | $b_0$ | $a_1b_1c_0 - a_1b_0c_0$ = (ac)-(a) |
| $a_0$ | $b_1$ | $a_0b_1c_1 - a_0b_1c_0$ = (bc)-(b) |
| $a_1$ | $b_1$ | $a_1b_1c_1 - a_1b_1c_0$ = (abc)-(ab) |

**Simple effect of C**

$$= \frac{1}{4}[(abc)-(ab)+(bc)-(b)+(ac)-(a)+(c)-(1)]$$

$$= \frac{1}{4}[ab(c-1)+b(c-1)+a(c-1)+1(c-1)]$$

$$= \frac{1}{4}[(c-1)(ab+b+a+1)]$$

$$= \frac{1}{4}[(c-1)\{b(a+1)+1(a+1)\}]$$

$$= \frac{1}{4}[(c-1)\{(a+1)(b+1)\}]$$

$$= \frac{1}{4}[(c-1)(a+1)(b+1)]$$

S.E. of C = $\frac{1}{4}[(a+1)(b+1)(c-1)]$

**Interaction effect of AB:**

AB = $\frac{1}{4}[\{(abc)-(bc)+(ab)-(b)\}-\{(ac)-(c)+(a)-(1)\}]$

$$= \frac{1}{4}[(abc)-(bc)+(ab)-(b)-(ac)+(c)-(a)+(1)]$$

(In main effect A we are add $b_1$ terms and substact $b_0$ terms)

$$= \frac{1}{4}[bc(a-1)+b(a-1)-c(a-1)-1(a-1)]$$
$$= \frac{1}{4}[(a-1)(bc+b-c-1)]$$
$$= \frac{1}{4}[(a-1)\{b(c+1)-1(c+1)\}]$$
$$= \frac{1}{4}[(a-1)(b-1)(c+1)]$$

I.E of B $= \frac{1}{4}[(a-1)(b-1)(c+1)]$

**Interaction effect of BC:**

$$\text{AC} = \frac{1}{4}[\{(abc)-(bc)+(ab)-(b)\}-\{(ac)-(c)+(a)-(1)\}]$$

(In main effect A we are add $c_1$ terms and subtract $c_0$ terms)

$$= \frac{1}{4}[(abc)-(bc)+(ac)-(c)-(ab)+(b)-(a)+(1)]$$
$$= \frac{1}{4}[bc(a-1)+c(a-1)-b(a-1)-1(a-1)]$$
$$= \frac{1}{4}[(a-1)(bc+c-b-1)]$$
$$= \frac{1}{4}[(a-1)\{c(b+1)-1(b+1)\}]$$
$$= \frac{1}{4}[(a-1)(b+1)(c-1)]$$

I.E. of AC $= \frac{1}{4}[(a-1)(b+1)(c-1)]$

**Interaction effect of BC**

$$\text{BC} = \frac{1}{4}[\{(abc)-(ac)+(bc)-(c)\}-\{(ab)-(a)+(b)-(1)\}]$$

(In main effect B we are add $c_1$ terms and subtract $c_0$ terms)

$$= \frac{1}{4}\left[(abc)-(ac)+(bc)-(c)-(ab)+(a)-(b)+(1)\right]$$
$$= \frac{1}{4}\left[ac(-1)+c(b-1)-a(b-1)-1(b-1)\right]$$
$$= \frac{1}{4}\left[(b-1)(ac+c-a-1)\right]$$
$$= \frac{1}{4}\left[(b-1)\{c(a+1)-1(a+1)\}\right]$$
$$= \frac{1}{4}\left[(b-1)(a+1)(c-1)\right]$$

Interaction effect of AC = $\frac{1}{4}\left[(a+1)(b-1)(c-1)\right]$

Interaction effect of ABC:

Interaction effect of AB at $C_0$ of C = $\frac{1}{2}\left[(a_1b_1-a_0b_1)c_0-(a_1b_0-a_0b_0)c_0\right]$

Interaction effect of AB at $C_1$ 0f C = $\frac{1}{2}\left[(a_1b_1-a_0b_1)c1-(a_1b_0-a_0b_0)c_1\right]$

Interaction effect of ABC = $\frac{1}{4}\left[(abc)-(bc)-(ac)+((c)-(ab)+(b)+(a)-(1))\right]$

$$= \frac{1}{4}\left[bc(a-1)-c(a-1)-b(a-1)+1(a-1)\right]$$
$$= \frac{1}{4}\left[(a-1)(bc-c-b+1)\right]$$
$$= \frac{1}{4}\left[(a-1)\{c(b-1)-1(a-1)\}\right]$$
$$= \frac{1}{4}\left[(a-1)(b-1)(c-1)\right]$$

I.E of ABC = $\frac{1}{4}\left[(a1)(b-1)(c-1)\right]$

### 4.3.1. Analysis of $2^3$ Factorial Design:

$$Y_{ijkl} = \mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\beta\alpha)_{jk}+(\alpha\gamma)_{ik}+(\alpha\beta\gamma)_{ijk}+\rho_l+\varepsilon_{ijkl}$$

Where, $Y_{ijkl}$ is the yield from the $i^{th}$ row $j^{th}$Block μ is the general mean effect.

$\alpha_i$ is the effect due to $i^{th}$ level of Treatment A.

$\beta_j$ is the effect due to $j^{th}$ level of Treatment B.

$\gamma_k$ is the effect due to $k^{th}$ level of Treatment C.

$(\alpha\beta)_{ij}$ is the interaction effect due to i and $j^{th}$ level of Treatment AB.

$(\beta\gamma)_{ij}$ is the interaction effect due to j and $k^{th}$level of Treatment BC.

$(\alpha\gamma)_{ik}$ is the interaction effect due to i and $k^{th}$ level of Treatment AC.

$(\alpha\beta\gamma)_{ijk}$ is the interaction effect due to i,j and $k^{th}$ level of Treatment ABC.

$\rho_l$is the effect of due to the $l^{th}$ Block.

$\varepsilon_{ijkl}$is the error effect.

The above parameter are subject to following restrictions

$\Sigma\alpha_i=0$, $\Sigma(\alpha\beta)_{ij}=0$

$\Sigma\beta_j=0$, $\Sigma(\beta\alpha)_{ij}=0$

$\Sigma(\alpha\beta\gamma)_{ijk}=0$, $\Sigma(a\gamma)_{ik}=0$

$\Sigma\rho_l=0$

**Null Hypothesis:**

$H_{OR}$: - there is no significance difference between replicates.

$H_{OT}$: - there is no significance difference between Treatment (or) Factorial not present.

**Statistical Analysis for $2^3$ factorial experiment**

The sum of square due to any factorial effect is obtain as multiplying the square of the effect total by the factor $\frac{1}{8r}$, where r is common replicate number.

Sum of square due to main effect A = $\frac{[A]^2}{8r}$

Sum of square due to main effect B = $\frac{[B]^2}{8r}$

Sum of square due to interaction effect AB = $\frac{[AB]^2}{8r}$

Sum of square due to interaction effect A = $\frac{[AC]^2}{8r}$

Sum of square due to interaction effect C = $\frac{[C]^2}{8r}$

Sum of square due to interaction effect BC = $\frac{[BC]^2}{8r}$

Sum of square due to interaction effect ABC = $\frac{[ABC]^2}{8r}$ each with 1 d.f

**Table 4.6: Yates method for $2^3$ factorial experiment: -**

| Treatment Combination | Treatment Total | (3) | (4) | (5) | Effect total (6) |
|---|---|---|---|---|---|
| 1 | [1] | [1]+[a] | [1]+[a]+[b]+[ab] | [1]+[a]+[b]+[ab]+[c]+[ac]+[bc+[abc] | G |
| a | [a] | [b]+[ab] | [c]+[ac]+[bc]+[abc] | [a]-[1]+[ab]-[b]+[ac]+[ac-[c]+[abc]-[bc] | [A] |
| b | [b] | [c]+[ac] | [a]-[1]+[ab]-[b] | [b]+[ab]-[1]-[a]+[bc]+[abc]-[c]-[ac] | [B] |
| ab | [ab] | [bc]+[abc] | [ac]-[c]+[abc]-[bc] | [ab]-[b]-[a]+[1]+[abc]-[bc]-[ac]+[c] | [AB] |
| c | [c] | [a]-[1] | [b]+[ab]-[1]-[a] | [c]+[ac]+[bc]+[abc]-[1]-[a]-[b]-[ab] | [C] |
| ac | [ac] | [ab]-[b] | [bc]+[abc]-[c]-[ac] | [ac]-[c]+[abc]-[bc]-[a]+[1]-[ab]+[b] | [AC] |

| | | | | | |
|---|---|---|---|---|---|
| bc<br><br>abc | [bc]<br><br>[abc] | [ac]-[c]<br><br>[abc]-[bc] | [ab]-[b]-[a]+[1]<br><br>[abc]-[bc]-[ac]+[c] | [bc]+[abc]-[c]-[ac]-[b]-[ab]+[1]+[a]<br><br>[abc]-[bc]-[ac]+[c]-[ab]+[b]+[a]-[1] | [BC]<br><br>[ABC] |

| **Sum of variance** | **Sum of square** | **d.f** | **Mean sum of square** | **F- ratio** |
|---|---|---|---|---|
| Replication | $S_R^2$ | r-1 | $\frac{S_R^2}{r-1} = S_R^2$ | $\frac{S_R^2}{S_E^2} = F_R$ |
| Treatment | $S_t^2$ | 7 | $\frac{S_t^2}{7} = S_t^2$ | $\frac{S_t^2}{S_E^2} = F_t$ |
| Main effect A | $S_A^2$ | 1 | $\frac{S_A^2}{1} = S_A^2$ | $\frac{S_A^2}{S_E^2} = F_A$ |
| Main effect B | $S_B^2$ | 1 | $\frac{S_B^2}{1} = S_B^2$ | $\frac{S_B^2}{S_E^2} = F_B$ |
| Int. effect AB | $S_{AB}^2$ | 1 | $\frac{S_{AB}^2}{1} = S_{AB}^2$ | $\frac{S_{AB}^2}{S_E^2} = F_{AB}$ |
| main effect C | $S_C^2$ | 1 | $\frac{S_C^2}{1} = S_C^2$ | $\frac{S_C^2}{S_E^2} = F_C$ |
| Int. Effect AC | $S_{AC}^2$ | 1 | $\frac{S_{AC}^2}{1} = S_{AC}^2$ | $\frac{S_{AC}^2}{S_E^2} = F_{AC}$ |
| Int. Effect BC | $S_{BC}^2$ | 1 | $\frac{S_{BC}^2}{r-1} = S_{BC}^2$ | $\frac{S_{BC}^2}{S_E^2} = R_{BC}$ |

| | | | | |
|---|---|---|---|---|
| Int. Effect ABC | $S_{ABC}^2$ | 1 | $\frac{S_{ABC}^2}{r-1} = S_{ABC}^2$ | $\frac{S_{ABC}^2}{S_E^2} = R_{ABC}$ |
| Error | $S_E^2$ | 7r-1) | $\frac{S_E^2}{7(r-1)} = S_E^2$ | - |
| Total | $S_T^2$ | 8r-1 | - | - |

**Table value:**

F [(r-1), (7(r-1))] at 5% level = $F_1$

F [7, 7(r-1)] at 5% level = $F_2$

F [1, 7(r-1)] at 5% level = $F_3$

**Conclusion:**

(i) $F_R < F_1$ we need not reject the null hypothesis $H_0$ and we conclude that there is no significance difference between replicates.

(ii) $F_t < F_2$ we need not reject the null hypothesis $H_0$ and we conclude that there is no significance difference between Treatments.

(iii) $F_A < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in main effect A.

(iv) $F_B < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in main effect B

(v) $F_{AB} < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in interaction effect AB.

(vi) $F_C < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in main effect C.

(vii) $F_{AC} < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in interaction effect AC.

(viii) $F_{BC} < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in interaction BC.

(ix)$F_{ABC} < F_3$ we need not reject the null hypothesis $H_0$ and we conclude that there is no Factorial effect in interaction effect ABC.

### 4.4. $2^n$ – Factorial Experiment

The results and the notations of $2^2$ and $2^3$ can be generalized to the case of $2^n$ experiment. g Here we consider n factors each at 2 levels. Suppose A, B, C, D,…, K are the factors each at two levels(0, 1). Corresponding small letters a, b, c, d,…,k denote the corresponding factors at the second level, the first level of any factor being signified by the absence of the corresponding small letter. The treatment combinations, in standard order can be written as:

1, a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd, etc.

For $2^n$-experiment, the various factorial effects are enumerated as follows:

Main effects : $n_{C_1}$ in number

Two-factor interactions : $n_{C_2}$ in number

Three –factor Interactions : $n_{C_3}$ in number

.

.

.

n factor interaction : $n_{C_n}$ in number

Hence, the total number of factorial effects in 2n –experiment are

$^nC_1+^nC_2+\ldots+^nC_n=[^nC_0+^nC_1+\ldots+^nC_n]-1$

$=(1+1)^n-1=2^n-1$

Main Effects and Interactions

As in the case of $2^2$ and $2^3$- experiment the results for the main effects and interactions can be generalized to the case $2^n$ experiment. Thus, for n factors A, B, C, D, …,K the main effects and interactions are given by the expression.

$$\frac{1}{2^{n+1}}[(a\pm1)(b\pm1)(c\pm1)(d\pm1)\cdots(k\pm1)]$$

### 4.4.1. Analysis of $2^n$ design

It will be seen that all the factorial effects (main and interaction) are mutually orthogonal contrasts of treatment totals. Hence, having obtained the factorial effect totals by Yates technique, the S.S due to each factorial effect is given by

$$\frac{[\ ]^2}{\sum_{i=1}^{2^n} r.1^2}=\frac{[\ ]^2}{r.2^n}$$

Where [ ] is the factorial effect total.

**Table 4.8. ANOVA Table**

| **Source of variation** | **Sum of square** | **d.f** | **Mean sum of square** |
|---|---|---|---|
| | $S_R^2=\frac{\sum B_j^2}{2^n}-C.F$ | r-1 | $\frac{S_R^2}{r-1}=s_R^2$ |
| Treatments | $S_T^2=\frac{\sum T_i^2}{r}-C.F$ | $2^n-1$ | $s_t^2=\frac{S_T^2}{2^n-1}$ |

| | | | |
|---|---|---|---|
| Main effect A | $S_A^2$=[A]$^2$/r.2$^n$ | 1 | $s_A^2 = S_A^2$ |
| Main effect B | $S_B^2$=[B]$^2$/r.2$^n$ | 1 | $s_B^2 = S_B^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ |
| Main effect K | $S_K^2$=[K]$^2$/r.2$^n$ | 1 | $s_K^2 = S_K^2$ |
| Two-factor Interactions | | | |
| AB | $S_{AB}^2$ = [AB]$^2$/r.2$^n$ | 1 | $s_{AB}^2 = S_{AB}^2$ |
| AC | $S_{AC}^2$ = [AC]$^2$/r.2$^n$ | 1 | $s_{AC}^2 = S_{AC}^2$ |
| BC | $S_{BC}^2$ = [BC]$^2$/r.2$^n$ | 1 | $s_{BC}^2 = S_{BC}^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ |
| Three-factor Interactions | | | |
| ABC | $S_{ABC}^2$ = [ABC]$^2$/r.2$^n$ | 1 | $s_{ABC}^2 = S_{ABC}^2$ |
| ACD | $S_{ACD}^2$ = [ACD]$^2$/r.2$^n$ | 1 | $s_{ACD}^2 = S_{ACD}^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ |

| | | | |
|---|---|---|---|
| n-factor interactions<br><br>ABCD…K | $S^2_{AB\cdots K}$ = $[AB\ldots K]^2/r.2^n$ | 1 | $s^2_{AB\ldots K} = S^2_{AB\ldots K}$ |
| Error | $S^2_E$ = By subtraction | $(r-1)(2^n-1)$ | $s^2_E = \frac{S^2_E}{(r-1)(2^n-1)}$ |
| Total | Raw.S.S- C.F | $r.2^{n-1}$ | |

The block effects and the factorial effects ( main and interactions) can be tested for significance by comparing their means S.S with error S.S.

**UNIT- V**

**5.1 Confounding**

When only the portions of treatment combinations are allotted to block within a replication, the comparison between blocks in a replication represents some treatment comparison, either a main effect or on interaction. In such cases it is not possible to distinguish treatment comparisons from block comparisons. Such a mix up is termed as confounding.

**Advantages and disadvantages of Confounding**

The only and the greatest advantage of confounding scheme lies in the fact that it reduces the experimental error considerably by stratifying the experimental material into homogeneous sub sets or sub groups. The removal of the variation among incomplete blocks within replicates often results in smaller error mean square as compared with a randomised complete block design, thus making the comparisons among some treatments more precise.

The following are the disadvantages of confounding

1. The confounded contrasts are replicated fewer times than are the other contrasts and as such there is loss of information on them and they can be estimated with a lower degree of precisions as the number of replications for them is reduced.

2. The algebraic calculations are usually more difficult and the statistical analysis is complex, specifically when some of the units are missing.
3. A number of problems arise if the treatments interact with blocks.

**5.2 Partial and complete confounding**

The same interactions are confounding in each replication (or) different sets of interaction are confounded in different replications. Both the procedures are practised if the same set of the interactions is confounded in all the replications confounding are called complete confounding.

If again different sets of interaction are confounded in different replications confounding is called partial confounding

**5.3 Complete confounding $2^3$ experiments**

In a $2^3$ experiment the 8 treatment combination require 8 units of homogeneous material each two from a block.

For example:Let us consider confounding the highest order interactions ABC we know that the interaction effect ABC is given by

$$\text{ABC} = \frac{1}{4}\left[(abc)-(bc)-(ac)+(c)-(ab)+(b)+(a)-(1)\right]$$

$$= \frac{1}{4}\left[(abc)+(a)+(b)+(c)-(ab)-(bc)-(ac)-(1)\right]$$

ABC confounded with blocks.

Block 1: (1) (ab) (ac) (bc)

Replicate

Block 2: (a) (b) (c) (abc)

**Yates method for a $2^3$ experiment:**

Black size is four in the question

Block 1 1 (ab)

Block 2 (ac) (bc)

Block 3 (a) (b)

Block 4 (c) (abc)

Table 5.1

| **Treatment combination** | **Treatment Total** | **(3)** | **(4)** | **(5)** | **Effects total** |
|---|---|---|---|---|---|
| 1 | [1] | $[1]+[a]=u_1$ | $u_1+u_2=v_1$ | $v_1+v_2=w_1$ | G.T |
| a | [a] | $[b]+[ab]=u_2$ | $u_3+u_4=v_2$ | $v_3+v_4=w_2$ | [A] |
| b | [b] | $[c]+[ac]=u_3$ | $u_5+u_6=v_3$ | $v_5+v_6=w_3$ | [B] |
| ab | [ab] | $[bc]+[abc]=u_4$ | $u_7+u_8=v_4$ | $v_7+v_8=w_4$ | [AB] |
| c | [c] | $[a]-[1]=u_5$ | $u_2-u_1=v_5$ | $v_2-v_1=w_5$ | [C] |
| ac | [ac] | $[ab]-[b]=u_6$ | $u_4-u_3=v_6$ | $v_4-v_3=w_6$ | [AC] |
| bc | [bc] | $[ac]-[c]=u_7$ | $u_6-u_5=v_7$ | $v_6-v_5=w_7$ | [BC] |
| abc | [abc] | $[abc]-[bc]=u_8$ | $u_8-u_7=v_8$ | $v_8-v_7=w_8$ | Not estimatle |

This confound component contain in the (2r-1) d.f. The ANOVA table will be as follows.

**Table 5.2 : ANOVA Table:**

| S.V | S.S | d.f | M.S.S | F- ratio |
|---|---|---|---|---|
| Blocks | $S_b^2$ | 2r-1 | $S_b^2$ | $\frac{S_b^2}{S_E^2}=F_1$ |
| Treatment | $S_t^2$ | 6 | $S_t^2$ | $\frac{S_t^2}{S_E^2}=F_2$ |

| | | | | |
|---|---|---|---|---|
| A | $S_A^2$ | 1 | $S_A^2$ | $\frac{S_A^2}{S_E^2} = F_3$ |
| B | $S_B^2$ | 1 | $S_B^2$ | $\frac{S_B^2}{S_E^2} = F_4$ |
| AB | $S_{AB}^2$ | 1 | $S_{AB}^2$ | $\frac{S_{AB}^2}{S_E^2} = F_5$ |
| C | $S_C^2$ | 1 | $S_C^2$ | $\frac{S_C^2}{S_E^2} = F_6$ |
| AC | $S_{AC}^2$ | 1 | $S_{AC}^2$ | $\frac{S_{AC}^2}{S_E^2} = F_7$ |
| BC | $S_{BC}^2$ | 1 | $S_{BC}^2$ | $\frac{S_{BC}^2}{S_E^2} = F_8$ |
| Error | $S_E^2$ | 6(r-1) | $S_E^2$ | - |
| Total | $S_T^2$ | 8r-1 | - | - |

**Null Hypothesis: -**

$H_0$: confounding is not effective

$H_1$: confounding is effective

**Inference:**

If the calculated F value is less than the tabulated $F^*$ value we accept $H_0$. Otherwise we reject $H_0$.

### 5.4 Analysis of $2^3$ partial confounding

The analysis of $2^3$ partially confounded design differs from that of the ordinary $2^3$ factorial experiment replicated 4 times only in the calculation of the partially confounded interactions. Each interaction being estimated only from the three replicates in which the given interaction is not confounded.

Analysis of $2^3$ partially confounded design with four replications and 'r' such replications. Let us suppose that a number of repetitions say 'r', of the above pattern or layout are performed such the positions of the replication, Blocks within replications and Blocks within Blocks are randomised then the structure of the Anova table will be as follows.

Table 5.3

| S.V | d.f | S.S |
|---|---|---|
| Blocks | 8r-1 | $\frac{1}{4}\Sigma(\text{total of Blocks})^2 - \frac{G^2}{32r}$ |
| Treatment | 7 | $S_A^2 + S_B^2 + S_C^2 + S_{AB}^2 + S_{BC}^2 + S_{AC}^2 + S_{ABC}^2 = S_t^2$ |
| A | 1 | $[A]^2/32\ r$ |
| B | 1 | $[B]^2/32\ r$ |
| C | 1 | $[C]^2/32\ r$ |
| AB | 1 | $[AB]^2/24\ r$ |

| | | |
|---|---|---|
| AC | 1 | $[AC]^2/24\ r$ |
| BC | 1 | $[BC]^2/24\ r$ |
| ABC | 1 | $[ABC]^2/24\ r$ |
| Error | 24r-7 | (By difference)= $S_E^2$ |
| Total | 32r-1 | $S_T^2$ = Total S.S |

**Calculation of sum of square due to confounded effects:**

It has already been explained that sum of square for confounded effects are to be obtained from those replications only in which the given effect is not confounded from practical point of view these sum of square can be obtain from the table of Yates method for all the 4 replications by appling sum adjusting factor (A.F) for any confounded effects is computed as follows

(i) Note the replication in which the given effect is confounded.

(ii) Note the sign of (1) in the corresponding algebraic expression of the effect to be confounded.

If the sign is positive then

A.F = [Total of the block containing (1) of replicate in which the

Effect is confounded] – [Total of the block not containing

(1) of the replicate in which the effect is confounded]

$A.F = T_1 - T_2$ (say)

If the sign is negative,

$A.F = T_2 - T_1$

This adjusting factor will be subtracted from the Factorial effects total of the confounded effects obtained from Yates's method for all replicates.

**5.5 Balanced Incomplete Block Designs(BIBD)**

If in a block the number of experimental units or plots is smaller than the number of treatments, then the block is said to be incomplete and a design constituted of such blocks is called an incomplete block design.

Balanced incomplete block designs which were developed for experiments in plant breeding and agriculture selection comparisons among pairs of treatments is made with equal precision.

**Definitions**

**Incomplete Block Design(I.B.D)**

An incomplete block design is one having v treatments and b blocks each of size k such that each of the treatments is replicated r times and each pair of treatments occurs once and only once in the same blocks v, b, r, and k are known as the parameters of the I.B.D

**Balanced Incomplete Block Design**

An arrangement of v treatments in b blocks of k plots each($k<v$) is known as BIBD, if

i) Each treatment occurs once and only once in r blocks and
ii) Each pair of treatments occurs together in λ blocks.
BIBD is used when all treatment comparisons are equally important as it ensures equal precisions of the estimates of all pairs of treatment effects.

**Parameters of BIBD**

The integers v,r,b,k and λ are called the parameters of the BIBD., where

V = number of varies or treatments, b=number of blocks

K=block size , r= number of replicates for each treatment

λ= number of blocks in which any pair of treatments occurs together or number of times any two treatments occur together in a block. The following parametric relations serve as a necessary condition for the existence of a BIBD.

i) $Vr = bk$ ii) $\lambda(v-1) = r(k-1)$ iii) $b \geq v$( Fisher's Inequality)

**Theorem 5.1**

$Vr = bk$

Proof:

Since there are v treatments each replicated r times, total number of plots in the design is vr. Further since there are b blocks each of size k, there are bk plots in all.

Hence vr=bk

Incidence Matrix: Associated with any design D is the incidence matrix N = ($n_{ij}$), (i = 1,2,…,v; j = 1,2,…,b), where $n_{ij}$ denotes the number of times the $i^{th}$ treatment occurs in the $j^{th}$ block. Thus by the definition of a BIBD

$$N = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1b} \\ n_{21} & n_{22} & \dots & n_{2b} \\ \vdots & \vdots & \vdots & \vdots \\ n_{v1} & n_{v2} & \dots & n_{vb} \end{bmatrix} \quad \dots(5.1)$$

Where $n_{ij}$ =1, if $i^{th}$ treatment occurs in the $j^{th}$ block …(5.2)

=0, otherwise

We also observe, by definition of BIBD

$$\sum_{j=1}^{b} n_{ij} = \sum_{j=1}^{b} n_{ij}^2 = r; (i = 1,2,\dots,v) \quad \dots (5.3)$$

$$\sum_{i=1}^{v} n_{ij} = \sum_{i=1}^{v} n_{ij}^2 = k; (j = 1,2,\dots,v) \quad \dots (5.4)$$

$$\sum_{j=1}^{b} n_{ij} n_{ij} = \lambda; (i \neq l = 1,2,..,v) \quad \dots(5.5)$$

Since $n_{ij}n_{lj}$=1 if and only if $i^{th}$ and $l^{th}$ treatments occur together in the $j^{th}$ block otherwise it is zero and they occur together in λ blocks

If N' denotes the transpose of N then

$$NN' = \begin{bmatrix} \sum_j n_{1j}^2 & \sum_j n_{1j}n_{2j} & \cdots & \sum_j n_{1j}n_{vj} \\ \sum_j n_{2j}n_{1j} & \sum_j n_{2j}^2 & \cdots & \sum_j n_{2j}n_{vj} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_j n_{vj}n_{1j} & \sum_j n_{vj}n_{2j} & \cdots & \sum_j n_{vj}^2 \end{bmatrix}$$

$$= \begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}_{v\times v} \qquad \text{.... (5.6)}$$

**Theorem 5.2** $\lambda(v-1)=r(k-1)$ …(5.7)

Proof

Let us denote by $E_{mn}$ the m×n matrix all of whose elements are unity. From (5.6), we get

$$NN'E_{v1} = \begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}_{v\times v} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{v\times 1}$$

$$= \left[r + \lambda(v-1)\right] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Also $=[r+\lambda(v-1)]E_{v1}$ …(5.8)

$$NN'E_{v1} = N(N'E_{v1})$$

From (5.4)

$$= N \begin{bmatrix} n_{11} & n_{21} & \cdots & n_{v1} \\ n_{12} & n_{22} & \cdots & n_{v2} \\ \vdots & \vdots & & \vdots \\ n_{1b} & n_{2b} & & n_{vb} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= N \begin{bmatrix} \sum_i n_{i1} \\ \sum_i n_{i2} \\ \vdots \\ \sum_i n_{ib} \end{bmatrix} \quad = \text{N } N \begin{bmatrix} k \\ k \\ \vdots \\ k \end{bmatrix}$$

From (5.3)

$$= \text{k} \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1b} \\ n_{21} & n_{22} & \cdots & n_{2b} \\ \vdots & \vdots & & \vdots \\ n_{v1} & n_{v2} & \cdots & n_{vb} \end{bmatrix}_{v\times b} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{b\times 1}$$

$$= k \begin{bmatrix} \sum_j n_{1j} \\ \sum_j n_{2j} \\ \vdots \\ \sum_j n_{vj} \end{bmatrix} = \text{k } k \begin{bmatrix} r \\ r \\ \vdots \\ r \end{bmatrix}_{v\times 1} \qquad \text{...(5. 9)}$$

$=kr\ E_{v1}$ ... (5.10)

From (5.9) and (5.10), we get

$[r+\lambda(v-1)]E_{v1} = kr\ E_{v1}$

$r+\lambda(v-1) = kr$

i.e., $\lambda(v-1) = r(k-1)$

**Theorem 5.3** b ≥v (Fisher's Inequality)

Proof. From (5.6) the determinant of the matrix $|NN'| = \begin{bmatrix} r & \lambda & \lambda & \dots & \lambda \\ \lambda & r & \lambda & \dots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda & \lambda & \lambda & \cdots & r \end{bmatrix}_{v\times v}$

Adding $2^{nd}$ , $3^{rd}$ , …, vth columns to the first column and taking [r+(v-1)] common from the first column, we get

$$NN' = [r+(v-1)\lambda]\begin{bmatrix} 1 & \lambda & \lambda & \cdots & \lambda \\ 1 & r & \lambda & \cdots & \lambda \\ 1 & \lambda & r & \cdots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda & \lambda & \cdots & r \end{bmatrix}$$

$$[r+(v-1)\lambda]\begin{bmatrix} 1 & \lambda & \lambda & \cdots & \lambda \\ 0 & (r-\lambda) & 0 & \cdots & 0 \\ 0 & 0 & (r-\lambda) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & (r-\lambda) \end{bmatrix}$$

(subtracting first row from the 2nd, 3rd, …, vth row)

$=[r+(v-1)\lambda(r-\lambda)^{v-1}$ expanding by first column)

$=rk(r-\lambda)^{v-1}$ using (5.6)

Thus $|NN'| \neq 0$, for if r = λ then from (5.6) we get

$$(v-1)=(k-1) \quad \Rightarrow \quad v=k$$

Indicating that the design reduces to randomized block design. Hence, $NN'$ is non singular and consequently $Rank(NN') = v$ …(5.11)

Since v is the order of matrix $NN'$.

$$Rank(NN') = Rank(N)$$

$\therefore$ $Rank(N) = v$ **from (5.11)** **… (5.12)**

But since N is a v × b matrix, it ranks can be at most b.

V= rank N ≤ b b ≥ v as desired. …(5.13)

**Theorem 5.4**

(i) r ≥ k
(ii) b ≥ v+r-k

proof

(i) we have vr = bk $\Rightarrow r = \frac{b}{v}k$

since b ≥ v, we get r ≥ k

(ii) we have v-k ≥ 0 and r-k ≥ 0

$$\therefore \quad (v-k)(r-k) \geq 0 \quad \Rightarrow \quad \left(\frac{v}{k}-1\right)(r-k) \geq 0$$

i.e., $\frac{v}{k}(r-k)-(r-k) \geq 0$

$$\therefore \frac{vr}{k}-v \geq r-k \qquad \Rightarrow b \geq v+r-k \qquad \left[\because vr = bk\right]$$

**Symmetric BIBD**

Definition. A BIBD is said to be symmetric if b = v and r = k.

**Theorem 5.5** In a symmetric BIBD, the number of treatments common between any two blocks is λ

**Proof.**

We have

$$NN' = \begin{bmatrix} r & \lambda & \lambda & \cdots & \lambda \\ \lambda & r & \lambda & \cdots & \lambda \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda & \lambda & \lambda & & r \end{bmatrix}$$

$$= \begin{bmatrix} r-\lambda & 0 & 0 & \cdots & 0 \\ 0 & r-\lambda & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & r-\lambda \end{bmatrix} + \begin{bmatrix} \lambda & \lambda & \cdots & \lambda \\ \lambda & \lambda & \cdots & \lambda \\ \vdots & \vdots & & \vdots \\ \lambda & \lambda & & \lambda \end{bmatrix}$$

$=(r-\lambda)I_v+\lambda E_{vv}$

Where Iv is a unit matrix of order v.

Also for a symmetric BIBD, we have

$$(NN')^{-1} = \left[ I_v - \frac{\lambda}{r^2} E_{vv} \right]$$

$$\Rightarrow (N')^{-1} N^{-1} = \frac{1}{r-\lambda}\left[ I_v - \frac{\lambda}{r^2} E_{vv} \right]$$

Premultiplying by (N'), we get $N^{-1} = \frac{1}{r-\lambda}\left[N' - \frac{\lambda}{r^2} N'E_{vv}\right]$ …(5.14)

But it can be easily verified that for symmetric BIBD

N'Evv=NEvv=rEvv=k Evv

$$\frac{N'}{r} E_{vv} = E_{vv}$$

Substituting in (5.14), we get $N^{-1} = \frac{1}{r-\lambda}\left[N' - \frac{\lambda}{r} E_{vv}\right]$

Post multiplying by N, we have

$$I_v = \frac{1}{r-\lambda}\left[N'N - \frac{\lambda}{r} NE_{vv}\right] = \frac{1}{r-\lambda}\left[N'N - \lambda E_{vv}\right]$$

$N'N=(r-\lambda)I_v+\lambda E_{vv}$ …(5.15)

From (5.14) and (5.15), we get for a symmetric BIBD

N'N=N'N

Thus, the inner product of any two rows of N is equal to the inner product of any two columns of N, i.e., λ.

Hence, in case of a symmetric BIBD, any two blocks have λ treatments in common.

Reference Books

Das, M.N. and Giri, N.C. (1988) Design and Analysis of Experiments( 2nd Edition). New Age International, New Delhi.

Douglas,C. Montgomenry (2012). Design and Analysis of Experiments. John Wiley and sons, New York.

Gupta, S.C. and V.K. Kapoor(1999). Fundamentals of Applied Statistics (Third Edition), Sultan Chand and Sons, New Delhi.