DESIGN OF EXPERIMENTS

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

In the words of Allen L. Edwards, "*The experimental design is called a randomised group design. The essential characteristic of this design is that subjects are randomly assigned to the experimental treatment or vice versa*."

Terminology in experimental designs (important terms and definitions)

A number of basic aspects used in the context of the theory of experimental design are worth noting at the very beginning.

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 (a) Absolute, and (b) Comparative.

Absolute experiments consists in determining the absolute value of some characteristics like:

(i) Obtaining the average intelligence quotient (I.Q.) of a group of people,
(ii) Finding the correlation coefficient between two variables in a bivariate distribution etc.

On the other hand, comparative experiments are designed to compare the effect of two or more objects on some population characteristics, e.g., comparison of different manures or fertilizers, different kinds of varieties of a crop, different cultivation processes, different pieces of land in a field experiment, or different diets or medicines in a dietary or medical experiment respectively.

Treatments. Various objects of comparison in a comparative experiment are termed as treatments, e.g., in field experimentation different fertilizers or different varieties of crop or different methods of cultivation are the treatments.

First of all, many experiments are conducted to establish the effect of one or more (independent) variables on a response (the dependent variable). Hear, the independent variables are often called treatments or factors which are often qualitative in nature, such as different makes of machines, different advertisement channels, different ways of packaging merchandise and so on.

The values of a response are supposed to reflect the effects of different treatments.

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, is termed as experimental unit, e.g., In field experiments the plot of land is the experimental unit. In other experiments, unit may be a patient in a hospital, a batch of seeds.

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

Yield. The measurement of the variable under study on different experimental units (e.g., Plots, between experiments) are termed as yields.

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 yield from some of the treatments are more than those of the 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 treatment is used on 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 (or chance or non-assignable) factors beyond human control, is spoken of as experimental error. It may be pointed out that the term error used here is not synonymous with mistake but is a technical term which includes are types of extraneous variations due to:

(i) The inherent variability in the experimental material to which treatments are applied,
(ii) The lack of uniformity in the methodology of conducting the experiment or in other words failure to standardise the experimental technique, and
(iii) Lack of representativeness of the sample to the population under study.

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

Precision. The reciprocal of the variance of the mean is termed as the precision, or the amount of information of a design. Thus, for an experiment replicated r times, the precision is given by:

$$\frac{1}{Var(\bar{x})} = \frac{r}{\sigma^2}$$

where σ^2 is the error variance per unit.

Efficiency of the 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:

$$\frac{2{\sigma_1}^2}{r_1} \; and \; \frac{2{\sigma_2}^2}{r_2} \; for \; D_1 \; and \; D_2 \; respectively. then \; the \; ratio$$

$$E = \frac{Var(D_2)}{Var(D_1)} = \frac{2{\sigma_2}^2}{r_2} \div \frac{2{\sigma_1}^2}{r_1}$$

is termed as efficiency of design D_1 with respect to D_2 . In other words, efficiency D_1 with respect to D_2 may be defined as the ratio of the positions of D_1 and D_2.

Three principles of experimental design

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

At the first instance replication serves to reduce experimental error and thus enables us to obtain more precise estimates of the treatment effects. From statistical theory we know that the standard error (S.E.) of the mean of a sample of size n is $\sigma/\sqrt{n}$, where σ is the standard deviation (per unit) of the population. Thus, if a treatment is replicated r times, then the S.E. of its mean effect is $\sigma/\sqrt{r}$, where σ^2 is the variance of the individual plot is estimated from the 'error variance'. Thus "the precision of the experiment is inversely

proportional to the square root of the replications". Consequently, replication has an important but limited role in increasing the efficiency of the design.

Randomisation. As discussed earlier, by replication the experimenter tries to average out as far as possible the effects due to uncontrolled factors. This brings to him the question of allocation of treatments to experimental units so that each treatment gets an equal chance of showing its worth. In the absence of the prior knowledge of the variability of experimental material, this objective is achieved through randomisation a process of assigning the treatments to various experimental units in a purely chance manner. The following are the main objectives of randomisation:

The purpose of randomness is to ensure that the sources of variation, not controlled in the experiment, operate randomly so that average effect on any group of units is zero. In other words, randomisation ensures that different treatments, by the repetition of the experiment, on the average are subject to equal environmental effect. Randomisation eliminates bias in any form. It equalizers even factors of variation over which we have no control.

Local control. If the experimental material, say field for agriculture experimentation, is heterogeneous and different treatments are allocated to various units (plots) at random over the entire field, the soil heterogeneity will also enter the uncontrolled factors and thus increase the experimental error. It is desirable to reduce the experimental error as far as practicable without unduly increasing the number of replications or without interfering with the statistical requirement of randomness, so that even smaller differences between treatments can be detected as significant. In addition to the principles of replication and randomisation discussed earlier, the experimental error can further be reduced by making use of the fact that neighbouring areas in a field are relatively more homogeneous than those widely spread. In order to separate self the soil fertility effects from the experimental error, the whole experimental area (field) is divided into homogeneous groups (blocks) row wise or column wise (one way elimination of fertility e gradient) or both (elimination of fertility gradient in two perpendicular directions), according to the fertility gradient of the soil such that the variation within each block is minimum and between the blocks is maximum. The treatments are then allocated at random within each block. The process of reducing the experimental error by dividing the relatively heterogeneous experimental area

(field) into homogeneous blocks (due to physical contiguity as far as field experiments are concerned) is known as local control.

COMPLETELY RANDOMISED DESIGN

The simplest and most flexible design is the completely randomised design. 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. In addition the units should be processed in random order at a subsequent stages in the experiment where this order is likely to affect the results. Also in this design treatments are allocated at random to the experimental units over the entire experimental material.

Let we have v treatments, the i^{th} treatment be replicated r_i times so that $n = \sum_i r_i$. The treatments may be numbered arbitrary from 1 to v, and the experimental units from 1 to n. r_1 units select at random from the n units, using a table of random numbers may be allocated to the first treatment, r_2 units selected randomly from the remaining units to the second treatment, and so on.

Advantages and Disadvantages of C.R.D.

Advantages. C.R.D. has several advantages explained below:

(i) It is easy to layout the design.
(ii) It results in the maximum use of the experimental units since all the experimental material can be used.
(iii) It allows complete flexibility as any number of treatments and replicates may be used for stop 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 or if the experimental errors differ from treatment to treatment.
(v) The relative loss of information due to missing data is smaller in comparison with any other design and they do not pose any problem in carrying out the standard analysis of data.
(vi) 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 i.e., for experiments with smaller number of treatments.

Disadvantages.

(i) In certain circumstances, the design suffers from the disadvantages of being inherently less informative than other more sophisticated layouts. This usually happens in 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 receiving the other treatment, the whole variations among the experimental units is included in the residual variance. This makes the design less efficient and results in less sensitivity and detecting significant effects.

Applications

(i) Completely randomised design is most useful in laboratory technique and methodological studies, i.e., In physics chemistry or cookery in chemical and biological experimentation in some greenhouse studies, etc., Where either the experimental material is homogeneous or the intrinsic variability between units can be reduced.

(ii) C.R.D. is also recommended in in situations where an appreciable fraction of units is likely to be destroyed or fail to respond.

Statistical Analysis of C.R.D.

Statistical analysis of a C.R.D. is analogous to the ANOVA for a one – way classified data for fixed effect model, the linear model (assuming various effects to be additive) becomes

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad (i = 1,2,..,v:\ j = 1,2,..,r_i) \quad \ldots\ldots(1)$$

Where,

$y_{ij} =$ the yield or response form the jth unit receiving the ith treatment,

$\mu =$ the general mean effect,

$\tau_i =$ the effect due to ith treatment

Such that $\sum_{i=1}^{n} r_i\tau_i = 0$, and

$\varepsilon_{ij} =$ is the error effect due to chance such that ε_{ij} is $iid\ N(0, {\sigma_e}^2)$

The effects $\tau_i \; and \; \mu$ are estimated by the principal of least squares on minimising the error sum of squares given by

$$E = \sum_{i}^{v}\sum_{j}^{r_i} \varepsilon_{ij}^2 \; = \; \sum_{i}^{v}\sum_{j}^{r_i} \left(y_{ij} - \; \mu - \; \tau_i\right)^2 \; \ldots\ldots (2)$$

The normal equations for estimating $\tau_i \; and \; \mu$ are obtained by differentiating with respect to $\mu \; and \; \tau_i$. First differentiating wrt μ.

$$\frac{\partial E}{\partial \mu} = \; -2 \left(\sum_{i}\sum_{j} \left(y_{ij} - \; \mu - \; \tau_i\right) \right) = 0,$$

$$\sum_{i}\sum_{j} y_{ij} - N\mu - \sum_{i} r_i \tau_i = 0$$

$$\sum_{i}\sum_{j} y_{ij} = N\mu \qquad \left\{ \sum_{i} r_i \tau_i = 0 \right\}$$

$$\mu = \frac{1}{N} \sum_{i}\sum_{j} y_{ij}$$

$$\hat{\mu} = \; \bar{y}..$$

Differentiating (2) with respect to τ_i

$$\frac{\partial E}{\partial \tau_i} = -2 \sum_{j} \left(y_{ij} - \; \mu - \; \tau_i\right) = 0$$

$$\sum_{j} y_{ij} \; - \; r_i \, \mu - \; r_i \tau_i = 0$$

$$r_i \tau_i = \; \sum_{j} y_{ij} \; - \; r_i \, \mu$$

$$\tau_i = \frac{1}{r_i} \sum_{j} y_{ij} \; - \; \mu$$

$$\tau_i = \bar{y}_i. - \; \mu$$

$$\hat{\tau}_i = \bar{y}_i. - \; \bar{y}..$$

Putting these values of μ and τ_i in equation (1)

$$y_{ij} \; = \; \mu + \; \tau_i + \; \varepsilon_{ij}$$

$$y_{ij} = \bar{y}..+(\bar{y}_i.- \bar{y}..) + \varepsilon_{ij}$$

$$y_{ij} = \bar{y}..+\bar{y}_i.- \bar{y}..+ \varepsilon_{ij}$$

$$y_{ij} = \bar{y}_i.+ \varepsilon_{ij}$$

$$\varepsilon_{ij} = y_{ij} - \bar{y}_i.$$

Now we have the values of μ, τ_i and ε_{ij} so our mathematical model becomes

$$y_{ij} = \bar{y}..+(\bar{y}_i.- \bar{y}..) + \left(y_{ij} - \bar{y}_i.\right)$$

Transposing $\bar{y}..$ to the left, squaring both sides and summing over i and j

$$\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}..\right)^2 = \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left[(\bar{y}_i.- \bar{y}..) + \left(y_{ij} - \bar{y}_i.\right)\right]^2$$

$$= \sum_{i=1}^{v}\sum_{j=1}^{r_i}(\bar{y}_i.- \bar{y}..)^2 + \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right)^2 + \sum_{i=1}^{v}\sum_{j=1}^{r_i}(\bar{y}_i.- \bar{y}..)\left(y_{ij} - \bar{y}_i.\right)$$

$$= \sum_{i=1}^{v} r_i(\bar{y}_i.- \bar{y}..)^2 + \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right)^2 + \sum_{i=1}^{v}(\bar{y}_i.- \bar{y}..)\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right)$$

$\{but\ since \sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right) = 0,$
$the\ algebric\ sum\ of\ deviations\ of\ observations\ from\ their\ mean\ is\ zero.\}$

$$\sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}..\right)^2 = \sum_{i=1}^{v} r_i(\bar{y}_i.- \bar{y}..)^2 + \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right)^2$$

$$T.S.S. = S.S.T. + S.S.E.$$

Where,

$$T.S.S. = \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}..\right)^2$$

$$S.S.T. = \sum_{i=1}^{v} r_i(\bar{y}_i. - \bar{y}..)^2 = {S_T}^2 \ (say)$$

$$S.S.E. = \sum_{i=1}^{v}\sum_{j=1}^{r_i}\left(y_{ij} - \bar{y}_i.\right)^2 = {S_E}^2 \ (say)$$

ANOVA table for C.R.D.

Source of Variation	d.f.	S. S.	M.S.S.	Variance ratio (F)
Treatments	$v-1$	${S_T}^2$	${s_T}^2 = \frac{{S_T}^2}{n-1}$	$F = \frac{{s_T}^2}{{s_E}^2}$
Error	$n-v$	${S_E}^2$	${s_E}^2 = \frac{{S_E}^2}{n-v}$	$(v-1, n-1)$
Total	$n-1$	${S_T}^2 + {S_E}^2$		

Under the null hypothesis H_0: $\tau_1 = \tau_2 = \cdots = \tau_v$ against the alternative that all $\tau's$ are not equal, the test statistic.

$$F = \frac{{s_T}^2}{{s_E}^2} \sim F\ (v-1, n-1)$$

If $F_{cal} > F_{tab}\ (v-1, n-1)\ d.f.$ at 5% level of significance, the null hypothesis is rejected. And we conclude that treatments differ significantly.

Otherwise H_0 may be accepted.

Que 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				
A	**55**	**49**	**42**	**21**	**52**
B	**61**	**112**	**30**	**89**	**63**
C	**42**	**97**	**81**	**95**	**92**
D	**169**	**137**	**169**	**85**	**154**

Solu: Null Hypothesis $Ho = \tau_A = \tau_B = \tau_C = \tau_D$

i.e., the treatment effects are same. In other words, all the treatments are alike as regard their effect on increase in weight.

Alternative Hypothesis H_0:At least two of the $\tau_i's$ are different.

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	137	169	85	154	714
	Grand Total					1695

Raw Sum of Squares

$$R.S.S. = \sum_i \sum_j {y_{ij}}^2 = (55^2 + 49^2 + \ldots.. + 154^2) = 181445$$

$$C.F. = \frac{G^2}{N} = \frac{(1695)^2}{20} = 143651.25$$

Now, Total Sum of Squares

$$T.S.S. = R.S.S. - C.F.$$

$$T.S.S. = 181445 - 143651.25 = 37793.75$$

And, Treatment Sum of Square

$$S.S.T. = \sum_{i=1}^{k} \left(\frac{{T_i}^2}{n_i} \right) - C.F.$$

$$S.S.T. = \left(\frac{219^2}{5} + \frac{355^2}{5} + \frac{407^2}{5} + \frac{714^2}{5} \right) - 143651.25$$

$$S.S.T. = (9592.2 + 25205 + 33129.8 + 101959.2) - 143651.25$$

$$S.S.T. = 26234.95$$

Now, error sum of squares

$$S.S.E = T.S.S. - S.S.T$$

$$S.S.E = 37793.75 - 26234.95 = 11558.80$$

ANOVA table

Source of variation	Sum of Squares	$d.f.$	Mean sum of Squares	Variance Ratio 'F'
Treatment Error	26234.95 11558.80	3 16	8744.98 722.42	$F_T = \frac{8744.98}{722.42} = 12.105$
Total	37793.75	19		

Now the tabulated value for F at 5% level of significance for (3,16) degrees of freedom is 3.24

Here $F_{cal} > F_{tab} \quad i.e. \quad 12.105 > 3.24$

Therefore, we reject the null hypothesis and conclude that the treatments differ significantly.

Que 2. For testing the variety effect in a completely randomised experiment, the data (the yield of barley in grams) are as shown in the following table:

320 V_1	**340** V_4	**398** V_5	**360** V_4	**350** V_3	**372** V_2	**455** V_2	**417** V_3	**420** V_1	**358** V_5
400 V_3	**353** V_1	**334** V_5	**331** V_1	**358** V_4	**370** V_4	**340** V_5	**375** V_2	**320** V_5	**325** V_3
430 V_5	**358** V_1	**378** V_3	**395** V_4	**328** V_2	**383** V_2	**275** V_3	**375** V_4	**308** V_2	**400** V_1

Ans: V.R. = 0.072

RANDOMISED BLOCK DESIGN (R.B.D.)

If the whole of the experimental area is not homogeneous and the fertility gradient is only in one direction, then a simple method of controlling the variability of the experimental material consist in stratifying for grouping the whole area into relatively homogeneous strata or subgroups (or blocks or replicates, as they are called), perpendicular to the direction of the fertility gradient. Now if the treatments are applied at random to relatively homogeneous units within each strata or block and a replicated over all the blocks the design is a randomised block design (RBD).

In a CRD we do not resort to the grouping of the experimental site (space, material or time) and allocate the treatments at random to the experimental units. But in RBD treatments are allocated at random within the units of each stratum or block, i.e., randomisation is restricted. Also variation among blocks is removed from variation due to error. Hence if it is desired to control one source of variation by stratification, the experimenter should select the RBD rather than CRD.

Layout of RBD. let us consider five treatments A, B, C, D and E each replicated four times. We divide the whole experimental area into four relatively homogeneous strata or blocks and each block into 5 units or plots. Treatments are then allocated at random to the plots of a block, fresh randomisation being done for each block. A particular layout may be as follows

Block I	D	B	E	A	C
Block II	E	A	C	D	B
Block III	A	C	B	D	E
Block IV	B	A	D	C	E

Advantages of RBD. Chief advantages of RBD may be outlined as follows:

1. accuracy. This design has been shown to be more efficient or accurate than CRD 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.
2. Flexibility. In RBD no restrictions are placed on the number of treatments are 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.
3. Ease of analysis. Statistical analysis is simple and straightforward. Moreover the error of any treatment can be isolated and any number of treatment may be omitted from the analysis without complicating it.

Disadvantages of RBD.

1. RBD may give misleading results if blocks are not homogeneous.
2. RBD is not suitable for large number of treatments because in that case the block size will increase and it may not be possible to keep large blocks homogeneous.

3. If the data on more than two plots is missing, the statistical analysis becomes white tedious and complicated.

Statistical Analysis of R.B.D. for One observation per Experimental Unit. In this case our linear model becomes:

$$y_{ij} = \mu + \tau_i + b_j + \varepsilon_{ij} \qquad \ldots\ldots(1)$$

Where,

$y_{ij} =$ the response or the yield of the experimental unit receiving the i[th] treatment in the j[th] block.

$\mu =$ the general mean effect.

$\tau_i =$ the effect due to the i[th] treatment.

$b_j =$ the effect due to the j[th] block.

$\varepsilon_{ij} =$the error effect due to chance.

Also $\varepsilon_{ij} \sim iid\ N(0, \sigma_e^2)$,

Where $\mu, \tau_i's\ \ and\ \ b_j's$ are constants so that $\sum_{i=1}^{t} \tau_i = 0$ and $\sum_{j=1}^{r} b_j = 0$

Also, if we write

$$\sum_i \sum_j y_{ij} = y.. = \ G$$

$$\sum_j y_{ij} = \ y_i. = T_i$$

$$\sum_i y_{ij} = \ y_{.j} = B_j$$

Now The effects $\tau_i \ \ \mu\ and\ \ b_j$ are estimated by the principal of least squares on minimising the error sum of squares given by

$$E = \sum_i^t \sum_j^r \varepsilon_{ij}^2 \ = \sum_i^t \sum_j^r \left(y_{ij} - \mu - \tau_i - b_j \right)^2 \ \ldots\ldots(2)$$

Differentiating (2) with respect to μ, we get

$$\frac{\partial E}{\partial \mu} = -2 \sum_{i}^{t}\sum_{j}^{r}\left(y_{ij} - \mu - \tau_i - b_j\right) = 0$$

$$\sum_{i}^{t}\sum_{j}^{r} y_{ij} - \sum_{i}^{t}\sum_{j}^{r} \mu - \sum_{i}^{t}\sum_{j}^{r} \tau_i - \sum_{i}^{t}\sum_{j}^{r} b_j = 0$$

$$\sum_{i}^{t}\sum_{j}^{r} y_{ij} - N\mu - \sum_{i}^{t} r\,\tau_i - \sum_{j}^{r} t\, b_j = 0$$

$$\sum_{i}^{t}\sum_{j}^{r} y_{ij} - N\mu - r\sum_{i}^{t} \tau_i - t\sum_{j}^{r} b_j = 0$$

$$\sum_{i}^{t}\sum_{j}^{r} y_{ij} - N\mu = 0 \qquad \left\{\sum_{i=1}^{t} \tau_i = 0 \text{ and } \sum_{j=1}^{r} b_j = 0\right\}$$

$$N\mu = \sum_{i}^{t}\sum_{j}^{r} y_{ij}$$

$$\hat{\mu} = \frac{1}{N}\sum_{i}^{t}\sum_{j}^{r} y_{ij} = \bar{y}..$$

Differentiating (2) with respect to τ_i, we get

$$\frac{\partial E}{\partial \tau_i} = -2 \sum_{j}^{r}\left(y_{ij} - \mu - \tau_i - b_j\right) = 0$$

$$\sum_{j}^{r} y_{ij} - \sum_{j}^{r} \mu - \sum_{j}^{r} \tau_i - \sum_{j}^{r} b_j = 0$$

$$\sum_{j}^{r} y_{ij} - r\,\mu - r\,\tau_i = 0$$

$$r\,\tau_i = \sum_{j}^{r} y_{ij} - r\,\mu$$

$$\tau_i = \frac{1}{r}\sum_{j}^{r} y_{ij} - \mu$$

$$\hat{\tau}_i = \bar{y}_i. - \bar{y}..$$

Differentiating (2) with respect to b_j, we get

$$\frac{\partial E}{\partial b_j} = -2 \sum_i^t \left(y_{ij} - \mu - \tau_i - b_j \right) = 0$$

$$\sum_i^t y_{ij} - \sum_i^t \mu - \sum_i^t \tau_i - \sum_i^t b_j = 0$$

$$\sum_i^t y_{ij} - t\,\mu - t\,b_j = 0$$

$$t\,b_j = \sum_i^t y_{ij} - t\,\mu$$

$$b_j = \frac{1}{t}\sum_i^t y_{ij} - \mu$$

$$\hat{b}_j = \bar{y}._j - \bar{y}..$$

Putting these values in our linear model,

$$y_{ij} = \mu + \tau_i + b_j + \varepsilon_{ij}$$

$$y_{ij} = \bar{y}.. + (\bar{y}_i. - \bar{y}..) + \left(\bar{y}._j - \bar{y}..\right) + \varepsilon_{ij}$$

$$\varepsilon_{ij} = y_{ij} + \bar{y}_i. + \bar{y}._j - \bar{y}..$$

Now our linear model becomes

$$y_{ij} = \bar{y}.. + (\bar{y}_i. - \bar{y}..) + \left(\bar{y}._j - \bar{y}..\right) + \left(y_{ij} + \bar{y}_i. + \bar{y}._j - \bar{y}..\right)$$

Transposing $\bar{y}..$ to the left and squaring both sides and summing over i and j

$$\sum_{i=1}^t \sum_{j=1}^r \left(y_{ij} - \bar{y}..\right)^2$$

$$= \sum_{i=1}^t \sum_{j=1}^r \left[(\bar{y}_i. - \bar{y}..) + \left(\bar{y}._j - \bar{y}..\right) + \left(y_{ij} + \bar{y}_i. + \bar{y}._j - \bar{y}..\right)\right]^2$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{y}_{i}. - \bar{y}..)^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{y}_{.j} - \bar{y}..)^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} + \bar{y}_{i}. + \bar{y}_{.j} - \bar{y}..)^2$$

$$= r\sum_{i=1}^{t}(\bar{y}_{i}. - \bar{y}..)^2 + r\sum_{j=1}^{r}(\bar{y}_{.j} - \bar{y}..)^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} + \bar{y}_{i}. + \bar{y}_{.j} - \bar{y}..)^2$$

The product terms vanish because the algebraic sum of deviations from mean is zero. Thus

$$T.S.S. = S.S.T. + S.S.B. + S.S.E.$$

Where

$T.S.S. = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \bar{y}..)^2 =$ total sum of squares

$S.S.T. = S_T^2 = r\sum_{i=1}^{t}(\bar{y}_{i}. - \bar{y}..)^2 =$ treatment sum of squares

$S.S.B. = S_B^2 = r\sum_{j=1}^{r}(\bar{y}_{.j} - \bar{y}..)^2 =$ block (variety) sum of squares

$S.S.E. = S_E^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} + \bar{y}_{i}. + \bar{y}_{.j} - \bar{y}..)^2 =$ error sum of squares

ANOVA table for R.B.D.

Source of Variation	d.f.	S. S.	M.S.S.	Variance ratio (F)
Treatments	$t-1$	${S_T}^2$	${s_T}^2 = \frac{{S_T}^2}{t-1}$	$F_T = \frac{{s_T}^2}{{s_E}^2}$
Blocks or Replicates	$r-1$	S_B^2	${s_T}^2 = \frac{{S_B}^2}{r-1}$	$F_B = \frac{{s_B}^2}{{s_E}^2}$
Error	$(t-1)(r-1)$	${S_E}^2$	${s_E}^2 = \frac{{S_E}^2}{(t-1)(r-1)}$	
Total	$rt-1$			

Under the null hypothesis $H_{0t}: \tau_1 = \tau_2 = \cdots = \tau_t$ against the alternative that all $\tau's$ are not equal, the test statistic.

$$F_T = \frac{{s_T}^2}{{s_E}^2} \sim F\,[(t-1),(t-1)(r-1)]$$

If $F_{cal} > F_{tab}\ [(t-1),(t-1)(r-1)]\ d.f.$ at 5% level of significance, the null hypothesis is rejected. And we conclude that treatments differ significantly.

Otherwise H_{0t} may be accepted.

Similarly, Under the null hypothesis $H_{0b}: b_1 = b_2 = \cdots = b_r$ against the alternative that all $b's$ are not equal, the test statistic.

$$F_B = \frac{{s_B}^2}{{s_E}^2} \sim F\ [(r-1),(t-1)(r-1)]$$

If $F_{cal} > F_{tab}\ [(r-1),(t-1)(r-1)]\ d.f.$ at 5% level of significance, the null hypothesis is rejected. And we conclude that blocks differ significantly.

Otherwise H_{0b} may be accepted.

Estimation of Missing value in R.B.D. let the observation $y_{ij} = x\ (say)$ in the j^{th} block and receiving the i^{th} treatment be missing, as given in table.

		Treatments						**Total**
		1	**2**	....	***i***	....	***t***	
Blocks	**1**	y_{11}	y_{21}	....	y_{i1}	....	y_{t1}	$y_{\cdot 1}$
	2	y_{12}	y_{22}	....	y_{i2}	....	y_{t2}	$y_{\cdot 2}$
	....	....	....	....	....	....	....	....
	j	y_{1j}	y_{2j}	....	$y_{ij} = x$	....	y_{tj}	$y_{\cdot j}' + x$
	....	....	....	....	....	....	....	....
	r	y_{1r}	y_{2r}	....	y_{ir}	....	y_{tr}	$y_{\cdot 1}$
	Total	$y_1.$	$y_2.$	....	$y_i.' + x$	....	$y_t.$	$y..' + x$

Where,

$y_i.'$ = total of known observations getting i^{th} treatment,

$y._j'$ = total of known observations in j^{th} block, and

$y..'$ = total of all the known observations

$$T.S.S. = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 - C.F. \quad = \quad x^2 + constant\ wrt\ x - C.F.$$

$$S.S.T. = \frac{1}{r}\left[(y_i.' + x)^2 + constant\ wrt\ x\right] - C.F.$$

$$S.S.B. = \frac{1}{t}\left[\left(y.'_j + x\right)^2 + constant\ wrt\ x\right] - C.F.$$

$$C.F. = \frac{(y..' + x)^2}{rt}$$

And

$$S.S.E. = T.S.S. - S.S.B. - S.S.T.$$

$$S.S.E. = x^2 - \frac{(y_i.' + x)^2}{r} - \frac{\left(y.'_j + x\right)^2}{t} + \frac{(y..' + x)^2}{rt} + constant\ wrt\ x$$

We shall choose x such that E is minimum. For E to be minimum for variations in x, we must have

$$\frac{\partial E}{\partial x} = 2x - \frac{2}{r}(y_i.' + x) - \frac{2}{t}\left(y.'_j + x\right) + \frac{2}{rt}(y..' + x) = 0$$

$$x\left(2 - \frac{2}{r} - \frac{2}{t} + \frac{2}{rt}\right) = \frac{2}{r}\, y_i.' + \frac{2}{t}\, y.'_j - \frac{2}{rt}\, y..'$$

$$x\left(1 - \frac{1}{r} - \frac{1}{t} + \frac{1}{rt}\right) = \frac{1}{r}\, y_i.' + \frac{1}{t}\, y.'_j - \frac{1}{rt}\, y..'$$

$$\frac{x(rt - t - r + 1)}{rt} = \frac{ty_i.' + ry.'_j - y..'}{rt}$$

$$x = \frac{ty_i.' + ry.'_j - y..'}{(r-1)(t-1)}$$

Two Missing Observations. For two missing values, say, x, and y, let R_1 and R_2 be the total of known observations in the rows containing x and y respectively and C_1 and C_2 be the totals of known observations in the columns containing x

and y respectively and let S be the total of all the known observations as given in the table

			x		R_1
	y				R_2
	C_2		C_1		

now

$$E = x^2 + y^2 - \frac{1}{t}\left[(R_1 + x)^2 + (R_2 + y)^2\right] - \frac{1}{r}\left[(C_1 + x)^2 + (C_2 + y)^2\right] + \frac{1}{rt}(S + x + y)^2$$

For a minima of E subject to the variations in x and y, we must have

$$\frac{\partial E}{\partial x} = 2x - \frac{2}{t}(R_1 + x) - \frac{2}{r}(C_1 + x) + \frac{2}{rt}(S + x + y) = 0$$

$$x - \frac{1}{t}(R_1 + x) - \frac{1}{r}(C_1 + x) + \frac{1}{rt}(S + x + y) = 0$$

$$x\left(\frac{1}{t} - \frac{1}{r} + \frac{1}{rt}\right) = \frac{1}{t}R_1 + \frac{1}{r}C_1 - \frac{1}{rt}S - \frac{1}{rt}y$$

$$x\left(\frac{rt - r - t + 1}{rt}\right) = \frac{rR_1 + tC_1 - s - y}{rt}$$

$$x(r-1)(t-1) = rR_1 + tC_1 - s - y$$

$$x = \frac{rR_1 + tC_1 - s - y}{(r-1)(t-1)}$$

Similarly

$$\frac{\partial E}{\partial y} = 2y - \frac{2}{t}(R_2 + y) - \frac{2}{r}(C_2 + y) + \frac{1}{rt}(S + x + y)$$

$$y = \frac{rR_2 + tC_2 - S - x}{(r-1)(t-1)}$$

On solving these two equations we get our values of x and y.

Que 3. Consider the results given in the following table for an experiment involving six treatments in four randomised blocks. The treatments are indicated by numbers within parentheses.

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

Test whether the treatments differ significantly. Also, Obtain the efficiency of this design relative to its analysis layout as C.R.D.

Solu: **Null hypothesis**: $H_{0t}: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6$ and

$H_{0b}: b_1 = b_2 = b_3 = b_4$ i.e. all the treatments as well as all the blocks are homogeneous.

Alternative hypothesis:

H_{1t}: at least two of the $\tau_i's$ are different.

H_{1b}: at least two of the $b_j's$ are different.

Blocks	Treatments						Block total B_j	$\sum_{i=1}^{6} y_{ij}^2$
	(1)	(2)	(3)	(4)	(5)	(6)		
1	24.7	20.6	27.7	16.2	16.2	24.9	130.3	2946.63

2	27.3	28.8	22.7	15.0	17.0	22.5	133.3	3110.27
3	38.5	39.5	36.8	19.6	15.4	26.3	176.1	5709.75
4	28.5	31.0	34.9	14.1	17.7	22.6	148.8	4014.12
Treatment total T_i	119	119.9	122.1	64.9	66.3	96.3	**588.5**	**15780.77**

Here $G = 588.5 \quad and\ R.S.S. = 15780.77$

$$C.F. = \frac{G^2}{rt} = \frac{(588.5)^2}{24} = \frac{346332.25}{24} = 14430.51$$

Here,

$$T.S.S. = R.S.S. - C.F.$$

$$T.S.S. = 15780.77 - 14430.51 = 1350.26$$

Sum of squares due to treatments

$$S.S.T = \frac{1}{4}\sum_{i=1}^{6} T_i^2 - C.F.$$

$$S.S.T = \frac{1}{4}(119^2 + 119.9^2 + 122.1^2 + 64.9^2 + 66.3^2 + 96.3^2) - 14430.51$$

$$S.S.T = \frac{1}{4}(61326.81) - 14430.51$$

$$S.S.T = 15331.70 - 14430.51$$

$$S.S.T = 901.19$$

Sum of squares due to bocks

$$S.S.B = \frac{1}{6}\sum_{j=1}^{4} B_j^2 - C.F.$$

$$S.S.B = \frac{1}{6}\,(130.3^2 + 133.3^2 + 176.1^2 + 148.8^2) - 14430.51$$

$$S.S.B = \frac{1}{6}\,(87899.63) - 14430.51$$

$$S.S.B = 14649.93 - 14430.51$$

$$S.S.B = 219.42$$

Sum of squares due to error

$$S.S.E. = T.S.S. - S.S.T - S.S.B$$

$$S.S.E. = 1350.26 - 901.19 - 219.42$$

$$S.S.E. = 229.65$$

ANOVA TABLE

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Treatments	901.19	5	$s_t^2 = \frac{901.19}{5} = 180.23$	$F_t = \frac{180.23}{15.31} = 11.77$
Blocks	219.42	3	$s_b^2 = \frac{219.42}{3} = 73.14$	$F_B = \frac{73.14}{15.31} = 4.77$
Error	229.65	15	$s_E^2 = \frac{229.65}{15} = 15.31$	
Total	1350.26	23		

Tabulated value of F for (5,15) degrees of freedom for 5% level of significance is 2.90

Here since, for treatments

$$F_{cal} > F_{tab}$$

$$11.77 > 2.90$$

The calculated value of F is much greater than the tabulated value, therefore, we may reject the H_{0t} and conclude that all the treatments are not homogeneous.

Tabulated value of F for (3,15) degrees of freedom for 5% level of significance is 3.29

Since, for blocks

$$F_{cal} > F_{tab}$$

$$4.77 > 3.29$$

The calculated value of F is much greater than the tabulated value, therefore, we may reject the H_{0b} and conclude that all the blocks are not homogeneous.

Que 5. In the given table, the yield of 6 varieties in a 4 replicates experiment for which one value is missing. Estimate, the missing value and analyse the data.

Blocks	**Treatments**					
	1	**2**	**3**	**4**	**5**	**6**
1	**18.5**	**15.7**	**16.2**	**14.1**	**13.0**	**13.6**
2	**11.7**		**12.9**	**14.4**	**16.9**	**12.5**
3	**15.4**	**16.6**	**15.5**	**20.3**	**18.4**	**21.5**
4	**16.5**	**18.6**	**12.7**	**15.7**	**16.5**	**18.0**

Solu: let the missing observation be x

here we calculate

Blocks	Treatments						Block totals B_j
	1	2	3	4	5	6	
1	18.5	15.7	16.2	14.1	13.0	13.6	91.1
2	11.7	x	12.9	14.4	16.9	12.5	$68.4 + x$
3	15.4	16.6	15.5	20.3	18.4	21.5	107.7
4	16.5	18.6	12.7	15.7	16.5	18.0	98.0
Treatment total T_i	62.1	$50.9 + x$	57.3	64.5	64.8	65.6	$365.2 + x$
$\sum_{i=1}^{6} y_{ij}^2$	988.55	$868.01 + x^2$	830.39	1064.75	1065.42	1127.46	$5944.58 + x^2$

Here we have $y_i.'\ 50.9 \qquad y.'_j = 68.4 \qquad and \quad y..' = 365.2$

$$r = 4 \quad and\ t = 6$$

$$x = \frac{ty_i.' + ry.'_j - y..'}{(r-1)(t-1)}$$

$$x = \frac{6(50.9) + 4(68.4) - 365.2}{(4-1)(6-1)}$$

$$x = \frac{305.4 + 273.6 - 365.2}{15}$$

$$x = \frac{213.8}{15}$$

$$x = 14.25$$

Now

$$R.S.S. = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 = 5944.58 + x^2$$

$$= 5944.58 + (14.25)^2$$

$$= 6147.64$$

$$C.F. = \frac{G^2}{rt} = \frac{(365.2 + x)^2}{24} = \frac{(379.45)^2}{24} = 5999.26$$

Now

$$T.S.S. = R.S.S. - C.F.$$

$$T.S.S. = 6147.64 - 5999.26 = 148.38$$

Sum of squares due to treatments

$$S.S.T = \frac{1}{4}\sum_{i=1}^{6} T_i^2 - C.F.$$

$$S.S.T = \frac{1}{4}(62.1^2 + (50.9 + x)^2 + 57.3^2 + 64.5^2 + 64.8^2 + 65.6^2) - 5999.26$$

$$S.S.T = \frac{1}{4}(24046.87) - 5999.26$$

$$S.S.T = 6011.71 - 5999.26$$

$$S.S.T = 12.45$$

Sum of squares due to bocks

$$S.S.B = \frac{1}{6}\sum_{i=1}^{4} B_j^2 - C.F.$$

$$S.S.B = \frac{1}{6}\left(91.1^2 + (68.4 + x)^2 + 107.7^2 + 98^2\right) - 5999.26$$

$$S.S.B = \frac{1}{6}\left(36333.52\right) - 5999.26$$

$$S.S.B = 6055.58 - 5999.26$$

$$S.S.B = 56.32$$

Sum of squares due to error

$$S.S.E. = T.S.S. - S.S.T - S.S.B$$

$$S.S.E. = 148.38 - 12.45 - 56.32$$

$$S.S.E. = 79.61$$

Null Hypothesis

$$H_{0t}: \tau_1 = \tau_2 = . \ \ldots . = \tau_6$$

$$H_{0b}: b_1 = b_2 = . \ \ldots . = b_6$$

That is all the treatments as well as blocks are homogeneous.

Alternative Hypothesis

$$H_{1t}: at\ least\ two\ of\ the\ treatments\ are\ different$$

$$H_{1b}: at\ least\ two\ of\ the\ blocks\ are\ different$$

Here one degree of freedom is lost for the total sum of squares and consequently from error sum of squares due to the missing value from the given data.

ANOVA TABLE

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	Variance ratio
Treatments	12.45	5	$s_t^2 = \frac{12.45}{5} = 2.49$	$F_t = \frac{2.49}{5.68} = 0.43$

Blocks	56.32	3	$s_b^2 = \frac{56.32}{3} = 18.77$	$F_B = \frac{18.77}{5.68} = 3.30$
Error	79.61	14	$s_E^2 = \frac{79.61}{14} = 5.68$	
Total	148.38	22		

Tabulated value of F for (5,14) degrees of freedom for 5% level of significance is 2.95

Here since, for treatments

$$F_{cal} < F_{tab}$$

$$0.43 < 2.95$$

The calculated value of F is much less than the tabulated value, therefore, we may accept the H_{0t} and conclude that all the treatments are homogeneous.

Tabulated value of F for (3,14) degrees of freedom for 5% level of significance is 3.34

Since, for blocks

$$F_{cal} < F_{tab}$$

$$3.30 < 3.34$$

The calculated value of F is less than the tabulated value, therefore, we may accept the H_{0b} and conclude that all the blocks are homogeneous.

Efficiency of R.B.D. Relative to C.R.D.

Consider a design with t treatments, each replicated r times. Then ANOVA tables for R.B.D. is given in the following table

Source of variation	Degrees of freedom	Mean sum of squares
Blocks	$r-1$	s_b^2
Treatments	$t-1$	s_t^2
Error	$(r-1)(t-1)$	s_E^2
Total	$rt-1$	

If we apply uniformity to trials to R.B.D. which consists in using the same treatment on all the rt units, there are no treatment variations.

Hence, consequently the treatment d.f. add to the error d.f.

Now, error degrees of freedom (due to uniformity trials)

$$= (r-1)(t-1) + (t-1)$$

$$= rt - t - r + 1 + t - 1$$

$$= rt - r = r(t-1)$$

Therefore, the error sum of squares is given by

$$S_E^2 = r(t-1)s_E^2$$

Now if we carry out same experiment as C.R.D. (on the same set of rt experimental units), then there is no variation due to the blocks. Hence the block d.f. and the block sum of squares add to the error d.f. and error sum of squares respectively.

Hence, for C.R.D.

$$Error\ d.f. = r(t-1) + (r-1) = rt - 1$$

Error mean sum of square for C.R.D. is given by

$$s_E'^2 = \frac{r(t-1)s_E^2 + (r-1)s_B^2}{rt-1}$$

Now the efficiency E of R.B.D. relative to C.R.D. is given by:

$$E = \frac{s_E'^2}{s_E^2} = \frac{r(t-1)s_E^2 + (r-1)s_B^2}{(rt-1)s_E^2}$$

This is the expression for efficiency.