Combinatorics & DOE

The subject matter of DOE includes

1. Planning of experiment
2. Obtaining relevant information from it regarding the statistical ~~information~~ ~~regard~~ hypothesis understood.
3. Making a statistical analysis of the data

Analysis of variance (ANOVA)

It plays a central role in DOE because it provides the formal statistical framework to analyze the experimental data. & draw valid conclusion

We use t-test for two variables

Null : $\mu_1 = \mu_2$

Alt : $\mu_1 \neq \mu_2$

But when we have more than two variables = ANOVA

ANOVA divides (partitions) total variability in experimental data into

1. Treatment effects / factor effects.
2. Error (random uncontrolled variation)

→ This helps determine wheather observed differences are due to controlled factors or just random noise.

Test significance of factor effects :

ANOVA allows hypothesis testing such as

1. Two different treatments produced significalhtly different responses.
2. Is the factor or combinations of factors is statistically important.

✱ For ANOVA we use F-test

ANOVA provides estimates of treatment means and error variance estimates.

Terminology

1. Experiment : An experiment is a device of getting an answer to the problem under consideration. It can be classified into two categories.
   a. Absolute experiment : obtaining accurate values.
   b. Comparitive experiment:
   - Consist in determining the absolute value of some charecteristic, like obtaining the IQ of two groups, find the coefficients
   - Designed to compare the effect of two or more objects on some population characteristics, eg : effect of different fertilizers.

2. Treatment : the various objects of comparisson in a comparitive experiment are termed as treatments. eg. different fertilizers are different dozes of treatments.

3. Experimental unit : smallest division of the experimental material to which we apply the treatments or on which we make observations on the variable study is known as the experimental unit. eg. Patient in a hospital or a plot of land.

4 Experimental error: This refers to the variation in the observed response that can't be explained by the factors included in the experiments, it represents random uncontrolled or inherent variability in the experimental process.

Precision:

$$\frac{1}{Var(\bar{x})} = \frac{8}{\sigma^2} \qquad E = r_1 / \sigma_1^2$$

Principles of n experimental design or DOE:
this is given by R.A. Fisher

i) Replication
ii) Randomization
iii) Local control

i) Replication means the repitition of treatments under investigation, the repetition provides more reliable and precise estimates then in possible in single observation By replication we are trying to average out we are trying to average out as far as possible the effects due to uncontrolled variables.

ii) Randomization means the allocation in random. It's the process of randomly assigning treatments to experimental units under randomization the allocation of treatments is such that each treatment gets an equal chance of showing it's worth & it eliminates bias.

Note: Randomization without replications is not of treatments.

sufficient its only when randomization to various units accompanied by an adequate no. of replication.

iii) Local control: the process of reducing the experimental error by dividing the relatively heterogeneous area into homogeneous groups is known as local controls. Within each block treatments are assigned randomly such that the variation within each block is minimum & between the blocks is maximum.

Completely Randomized design (CRD)

The completely Randomized " is the simplest of all the designs based on Replication & Randomization. In this design treatments are allocated at random to the experimental units over the entire experimental unit.

→ experimental material is homogenous.

draw-back. CRD can be used when the experimental units are homogenous.

Let us suppose that we have k treatments. & the i^{th} treatment being ~~replicated are i times. Where i is from 1 - k times~~ r_i $(i = 1, 2 \dots k)$ times

The whole experimental material is divided into

$$n = \sum_{i=1}^{k} r_i \quad \text{in experimental units.}$$

& treatments are distributed completely at random over the units subject to the condition that the i^{th} treatment occur r times.

Advantages of CRD.

1. CRD results in the maximum use of the experimental units, since all the exp. material

can be used.
2. This design is very flexible, any number of treatments can be used.
3. It provides the maximum number of degrees of freedom for the estimation of the error variance.

16/01
Lets revise the topic of CRD.

CRD : k treatments.

$$n = \sum_{n=1}^{k} n_j \qquad (n_j = 1, 2 \ldots k).$$

i) n_1 is taken out of n & assign T_1 to n_1.

ii) n_2 out of $(n - n_1)$ & assign T_2 to n_2.

eg) 4 treatments & 20

$$n = \sum r_i \quad \text{experimental units.}$$

1) T_1 is replicated, 3 times.

2) T_2 is replicated say 5 times

3) T_3 is replicated, 6 times.

4) T_4 is replicated $(20 - (\ \ ))$ 6 times.

Analysis of CRD
There is only one factor which is affecting outcome that is treatment effect, so for CRD the setup of one way analysis (One way ANOVA) is to be
one factor → Treatment

used.

Treatments

	1	2	....	...	k
1	y_{11}	y_{12}	--	---	y_{1k}
	y_{12}	y_{22}	--	-----	y_{2k}
	⋮	⋮			⋮
	y_{1n}	y_{2n}			y_{kn}
	T_1	T_2	--	----	T_k

$T_i = \sum_{j=1}^{n_i} y_{ij}$ $\qquad n = \sum_{i=1}^{k} n_i \quad ; \quad \sum n_i = n.$

where y_{ij} is the individual measurement of jth experimental units for ith treatments where $(i = 1, 2 \ldots k)$ & $(j = 1, 2 \ldots n_i)$

this means treatment

replication.

like $y_{21} \Rightarrow$ 2 treatment replicated 1 time

[In general the number of replications are equal in CRD] did not really understand.

Linear model for Analysis of CRD:

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad \text{--- (1)}$$

j or ϵ_{ij}

where μ = overall mean

α_i = ith treatment effect.

e_{ij} are independently & identitically (i.id) distributed random errors with

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

$$T_i = \sum_{j=1}^{n_i} y_{ij}$$

It is the treatment total due to ith treatment.

$$G = \sum_{i=1}^{k} T_i$$

Is the grand total of all the observation.

→ Null Hypothesis & alternative hypothesis to be tested against one way ANOVA.

$H_0 = \alpha_1 = \alpha_2 \dots = \alpha_k$

$H_1 =$ at least one $\alpha_i \neq 0$

→ Estimates of parameters of $y_{ij} = \mu + \alpha_i + e_{ij}$

we will use least square method over here.

Procedure

$$S = \sum_{i=1}^{k} \sum_{j=1}^{n_i} e_{ij}^2 = \sum_i \sum_j (y_{ij} - \mu - \alpha_i)^2$$

$\frac{\partial S}{\partial \mu} = 0$; $\frac{\partial S}{\partial \alpha_i} = 0$ } These are called normal eqns

$$\frac{\partial S}{\partial \mu} = -2 \sum_i \sum_j (y_{ij} - \mu - \alpha_i) = 0.$$

$$\Rightarrow \sum_i \sum_j (y_{ij} - \mu - \alpha_i) = 0. \quad \text{— (1)}$$

$$\frac{\partial S}{\partial \alpha_i} = -2 \sum_j (y_{ij} - \mu - \alpha_i) = 0$$

$$\Rightarrow \cancel{\sum_i} \sum_j (y_{ij} - \mu - \alpha_i) = 0. \quad \text{— (2)}$$

→ As it is already explained for i in $\partial S / \partial \alpha_i$ so only $\sum_j$ remains in the eqn (the index for i is already set).

Solving eqn (1)

$$\sum_i \sum_j y_{ij} - \sum_i \sum_j \mu - \sum_i^k \sum_j^{n_i} \alpha_i = 0$$

$$\sum_i \sum_j y_{ij} - n\mu - \sum_{i=1}^{k} n_i \alpha_i = 0.$$

One of the assumptions of one way classification of ANOVA is $\sum_i n_i \alpha_i = 0$.

$$\Rightarrow \sum_i \sum_j y_{ij} = n\mu$$

$$\hat{\mu} = \sum_i \sum_j y_{ij}$$

$$\boxed{\hat{\mu} = \bar{y}_{..}}$$

Solving eq^n (2) : $\sum_j y_{ij} - \sum_{j=1}^{n_i} \mu - \sum_j \alpha_i = 0$

$$\sum_j y_{ij} - n_i\mu - n_i\alpha_i = 0$$

$$\sum_j y_{ij} = n_i(\mu + \alpha_i)$$

$$\mu + \alpha_i = \frac{\sum_j y_{ij}}{n_i}$$

$$\hat{\alpha}_j = \frac{\sum_j y_{ij}}{n_j} - \hat{\mu}$$

$$- \bar{y}_{..}$$

$$\boxed{\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}}$$

overall mean

error $\sim N(0, \sigma^2)$

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

effect due to ith treatment

Analysis:

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0.$

$H_1 : \alpha_i{}^2 r \neq 0.$

20/01

One may classification of ANOVA

In one may classification of ANOVA.

Learn ANOVA it will help you with flow.

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.}) + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (1)$$

TSS = SSE + SST

total sum of squares; sum of squares; treatment

Proof: Here we have

$$\mu = \sum_{i=1}^{K} \frac{n_i \mu_i}{n} \; ; \; \boxed{\sum_i n_i \alpha_i = 0} \text{ ——— } (1a)$$

assumption

Testing

$$\begin{cases} H_0 : \alpha_1 = \alpha_2 \text{ ---- } = \alpha_K \\ H_1 : \end{cases}$$

$$\bar{y}_{i.} = \sum_j \frac{y_{ij}}{n_i} \; , \quad \bar{y}_{..} = \frac{\sum_{ij} y_i y_j}{n}$$

Mean of the ith treatment.

or

$$\bar{y}_{..} = \frac{\sum_i n_i \bar{y}_{i.}}{n}$$

Now take LHS of eq^n (1)

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j \left((y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})\right)^2$$

(add & subtract $\bar{y}_{i.}$)

$$\Rightarrow \sum_i \sum_j \left[(y_{ij} - \bar{y}_{i.})^2 + (\bar{y}_{i.} - \bar{y}_{..})^2 + 2 (y_{ij} - \bar{y}_i) \cdot (\bar{y}_{i.} - \bar{y}_{..}) \right]$$

$$a^2 + b^2 + 2ab$$

$$\rightarrow \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..}) + 2 \sum_j (\underset{\substack{\| \\ 0}}{y_{ij}} - \bar{y}_{..}) \sum_j (\bar{y}_{i.} - \bar{y}_{..})$$

Assummation of this minus mean is 0

$$LHS \Rightarrow \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i \sum_j (y_{i.} - \bar{y}_{..})^2$$

doubt

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \left(\sum_i \sum_j y_{ij} - \bar{y}_{i.}\right)^2 + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$TSS = SSE + SST$$

Total SS = SS due to error + SS due to treatment.

Source of variation	degrees of freedom	Sum of squares	Mean SS	F_{cal}	F value. F_{tab} or $F_{critical}$
Treatment	$k-1$	SST	$\frac{SST}{k-1} = MSST$	$= \frac{MSST}{MSSE}$	$F_{tab} = F_{\{(k-1)(n-k)\}}(\alpha^*)$
Error	$n-k$	SSE	$\frac{SSE}{n-k} = MSSE$		
Total	$n-1$	TSS	X		

α^* = Level of Significance

$F_{cal} > F_{tab} \Rightarrow$ Reject H_0 at $\alpha^*\%$.

For numerical computation

How to calculate

$$TSS = \underbrace{\sum_i \sum_j y_{ij}^2}_{\text{Raw Sum}} - \frac{G^2}{n}$$

n = total exp. units

G = grand total of all the observations

$\frac{G^2}{n}$ is called correction factor. (c.f)

$TSS = RSS - CF$

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

$$SST = \sum_{i=1}^{K} \frac{T_i^2}{n_i} - \frac{G^2}{n}$$

$$SSE = TSS - SST.$$

Q1) Analyse the CRD.

	Treatment.		
	A	B	C
1	23	42	47
2	36	26	43
3	31	47	43
4	33	34	39
	123	149	172

$G = 444.$

$H_0 : \alpha_A = \alpha_B = \alpha_C = 0$ → this means

$\alpha_A - \alpha_B = 0$
$\alpha_A - \alpha_C = 0$
& so on.

H_1: Alternate hypothesis

$$C.F = \frac{G^2}{n} = \frac{(444)^2}{12} = 16428$$

$$TSS = RSS - CF = 17108 - 16428$$
$$= 680.$$

$$SST = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} - C.F$$
$$= \frac{T_1^2 + T_2^2 + T_3^2}{4} - C.F$$
$$= 300.5$$

$SSE = 379.5$

ANOVA table

SOV	d.f	SS	MSS
Treatment	$3-1=2$	300.5	
Error	9	379.5	
Total	$12-1=11$	680	

21/01

Over here the null hypothesis is accepted $\hookrightarrow$ Significant.

But when.

Reject null hypothesis $\longrightarrow$ insignificant

Q) A set of data involving free stuff
A B C D

20 Chicks

All the 20 chicks are treated alike in all respects accept the feeding treatments & each feeding treatment is given to 5 chicks. Analyze the data using CRD.

Feeding stuff (treatment)

$n = 20$

A	B	C	D
55	61	42	169
49	112	97	137
42	30	81	169

21	89	95	85	
52	63	92	154	
219	855	407	687	1668 = G

$$C.F = \frac{1668}{20} = 83.4$$

$$TSS = RSS - CF$$
$$= 181445 - 83.4$$
$$= 181{,}361.6$$

$$SST = \frac{T_1^2 + T_2^2 + T_3^2 + T_4^2}{n} - C.F$$
$$= \frac{811604}{5} - 83.4$$
$$= 162{,}237.4$$

$$SSE = TSS - SST = 19{,}124.2$$

ANOVA table

SOV	d.f	SS	MSS	F-value F_{cal}	F-value F_{tab}
Treatment	$4-1=3$	162,237.4	$\frac{SST}{K-1} = 54079.13 = MSST$	45.2446	$F_{0.05(3,16)} = 3.06$
Error	20-4 16	19,124.2	$\frac{SSE}{n-k} = 1195.26 = MSSE$		
Total	19	181361.6	x		

$$C.D = \sqrt{\frac{2.MSSE}{r}} \times t_{(\alpha^*, \text{error df})}.$$

r = replication number.

$$= \sqrt{\frac{2 \times 722.425}{5}} \times 2.12$$

$$\boxed{CD = 36.018}$$

$$\left| \frac{T_1}{n_1} - \frac{T_2}{n_2} \right| > CD$$

$$\left| \frac{219}{5} - \frac{315}{5} \right| = 19.2 < CD$$

$\therefore$

Note

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 \dots = 0$
H_1 : All α's are not equal.

Reject $H_0 \Rightarrow$ Significant.
Why Rejection of H_0 is called Significant as there is a significant difference between the α's.

eg $\underline{\underline{AC}} > C.D$ then $A - C \neq 0$.
$\hookrightarrow$ then reject null hypothesis.

RBD - Randomized Block design.

- If a large number of treatments are to be compared then a large number of experimental units are required. This will increase the variation among the response variable & CRD may not be appropriate to use.
- In such a case when the experimental material is not homogenous & there are k treatments to be compared then it may be possible to group the experimental materials into blocks of size k units
- Blocks are constructed such that the experimental unit within a block are relatively homogenous & resemble to each other more closely than the units in differ. blocks.
  1. If there are b such blocks we say that the blocks are at b levels, similarly if there are k treatments we say that treatments are at k levels.
- The responses from the b levels of blocks & k levels

of treatments can be arranged in a two way layout & the observed dataset is arranged as follows.

Blocks (Factor A)	Treatments (Factor B) 1	2	...	k	Block Totals
1	y_{11}	y_{12}	...	y_{1k}	B_1
2	y_{21}	y_{22}	...	y_{2k}	B_2
⋮	⋮	⋮			⋮
b	y_{b1}	y_{b2}	...	y_{bk}	B_K
Treatment totals	T_1	T_2		T_k	G = Grant totals

23/01

Remarks

Since all the treatments are to be applied within each block in each block we take as many units as the number of treatments. Thus in RBD the number of blocks & the number of blocks & the number of replication are same.

no. of treatments = no. of plots $(t = k)$

plots $\rightarrow$ units of exp. material.

no. of blocks = no. of replication $(r = b)$.

eg 1) Suppose there are 7 treatments $T_1\ T_2 --- T_7$ corresponding to 7 levels of a factor. to be included in 4 blocks. So one possible layout of the assignment of 7 treatments to 4 blocks in RBD is as follows.

Block 1	T_2	T_7	T_3	T_5	T_1	T_4	T_6
Block 2	T_1	T_6	T_7	T_4	T_5	T_3	T_2
Block 3	T_7	T_5	T_1	T_6	T_4	T_2	T_3
Block 4	T_4	T_1	T_5	T_6	T_2	T_7	T_3

Possible Layout for RBD

$r = b$

$t - k$

no. of treatments

representation of treatment

* In RBD we have to affect one is block affect & other linear affect thus we use two way ANOVA.

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \rightarrow \text{error term} \sim N(0, \sigma^2)$$

overall mean effect

response variable

i^{th} treatment effect

j^{th} block effect.

Linear Model for RBD

Least square estimate

$$S = \sum_i \sum_j e_{ij}^2 = \sum_i \sum_j (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

three normal eqns as 3 parameters

$$\frac{\partial S}{\partial \mu} = 0, \quad \frac{\partial S}{\partial \alpha_i} = 0, \quad \frac{\partial S}{\partial \beta_j} = 0$$

$$\frac{\partial S}{\partial \mu} = \frac{\partial}{\partial \mu}\left[\sum_i \sum_j (y_{ij} - \mu - \alpha_i - \beta_j)^2\right]$$

$$= -2 \sum_i \sum_j (y_{ij} - \mu - \alpha_i - \beta_j) = 0$$

$$= \sum_i \sum_j (y_{ij} - \mu - \alpha_i - \beta_j) = 0$$

no. of treatmen

$n = rk$

no. of replication

$$= \sum_i \sum_j y_{ij} - rk\mu - \sum_{i=1}^{k} \sum_{j=1}^{r} \alpha_i - \sum_i \sum_j \beta_j = 0$$

$$= \sum_i \sum_j y_{ij} - rk\mu - r\sum_i \alpha_i - k\sum_j \beta_j = 0$$

$\therefore n\mu = rk\mu.$

$$\boxed{\sum_i \sum_j y_{ij} = rk\mu} \qquad \boxed{\sum_i \alpha_i = 0 = \sum_j \beta_j = 0}$$

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

$$\frac{\partial S}{\partial \alpha_i} = \sum_j (y_{ij} - \mu - \alpha_i - \beta_j)^2$$

$$= -2\sum_j (y_{ij} - \mu - \alpha_i - \beta_j) = 0.$$

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

$$= \sum_j y_{ij} - \sum_j \mu - \sum_j \alpha_i - \sum_j \beta_j$$

$$= \sum_j y_{ij} - r\mu - r\alpha_i = 0$$

$\sum_j y_{ij} = \sum T_i$

when we divide it by r it gives us the arithmetic mean of each treatment.

$$\sum_j y_{ij} = r(\mu + \alpha_i) \Rightarrow \frac{\sum_j y_{ij}}{r} - \mu = \alpha_i$$

$$\therefore \boxed{\bar{y}_{i.} - \bar{y}_{..} = \hat{\alpha}_i}$$

$$\frac{\partial S}{\partial \beta_j} = \sum_i \left(y_{ij} - \mu - \alpha_i - \beta_j\right)^2$$

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

$$= \sum_i y_{ij} - \sum_i \mu - \sum_i \alpha_i - \sum_i \beta_j = 0.$$

$$\sum_i y_{ij} - k\mu - 0 - k\beta_j = 0.$$

$$\frac{\sum_i y_{ij}}{k} = \mu + \beta_j \longrightarrow \boxed{\bar{y}_{.j} - \bar{y}_{..} = \beta_j}$$

Hypothesis testing

$$H_{0B} : \beta_1 = \beta_2 = \cdots = \beta_r = 0$$

$$H_{0T} : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

Alternative hypothe

$$H_B : \beta_1 \neq \beta_2 \cdots \neq \beta_r \neq 0$$

$$H_T : \alpha_1 \neq \alpha_2 \cdots \neq$$

Statistical Analysis of RBD

RBD is based on 2-may ~~analysis~~ classification of ANOVA & there are 2 null hypothesis to be tested.

① Related to Block effects.

② Related to treatment effect.

Fundamental of Fundamental statistic

27/01

$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$. $\longrightarrow$ Model of RBD.

$i = 1, 2 \ldots k$

$j = 1, 2 \ldots r$

y_{ij} = Response variable

μ = overall mean effect.

α_i = i^{th} treatment effect

β_j = j^{th} block effect.

e_{ij} = Error $\sim (0, \sigma^2)$

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

$$\sum_j y_{ij} = y_{i.}$$

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

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

$$+ \sum_i \sum_j (y_{ij} - \bar{y}_{.j} + \bar{y}_{..})^2 \quad \text{--- } (1).$$

$$TSS = SST + SSB + SSE$$

Take LHS of eq^n (1)

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j \left[y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} + \bar{y}_{.j} - \bar{y}_{..} \right]^2$$

add & subt ($\bar{y}_{i.}$, $\bar{y}_{.j}$ & $\bar{y}_{..}$)

$$= \sum_i \sum_j \left[\underset{a}{(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})} + \underset{b}{(\bar{y}_{i.} - \bar{y}_{..})} + \underset{c}{(\bar{y}_{.j} - \bar{y}_{..})} \right]^2$$

Product term vanished $2ab = 2ca = 2cb = 0$.

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

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = SSE + r \underbrace{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2}_{SST}$$
$$+ k \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

$$\boxed{TSS = SST + SSB + SSE}$$

For numerical computation the various sum of squares are:
Total SS i.e. $TSS = \sum_i \sum_j y_{ij} - C.F$ where $CF = \frac{G^2}{rk}$

$TSS = RSS - C.F.$
(raw sum of squares)

$SST =$ Treatment sum of squares. $= \sum_{i=1}^{k} \frac{T_i^2}{k} - C.F$

$$SSB = \sum_j \frac{B_j^2}{r} - C.F$$

$$SSE = TSS - SST - SSB.$$

ANOVA table for RBD.

Source of Variation	d/f	SS	MSS	F ratio: Fcal	F ratio: Ftab LOS
Treatment	$k-1$	SST	$MSST = \frac{SST}{k-1}$	$F_T = MSST/MSSE$	$F_{\{k-1, (r-1)(k-1)\}}(\alpha)$
Block	$r-1$	SSB	$MSSB = SSB/r-1$	$F_B = MSSB/MSSE$	$F_{\{(r-1), (r-1)(k-1)\}}$ at α%
Error	$(k-1)(r-1)$	SSE	$MSSE = SSE/(k-1)(r-1)$		

Total	$rk-1$	TSS		

Decision Making α% loss

If F_{cal} or $F_T > F_{\{(k-1),(r-1)(k-1)\}}$

then Reject H_{OT} or H_{OB} at α% LOS

$F_B > F_{\{(r-1),(r-1)(k-1)\}}$

Otherwise accept it.

Q1)

Blocks

Varities	I	II	III	IV
A	7	16	10	11
B	14	15	15	14
C	8	16	7	11

Perform statistical analysis of the given dataset in RBD

Yes I got angry but I calmed down & spoke to you properly instead you were just rude as usual.

Blocks

Varities	I	II	III	IV	Total
A	7	16	10	11	T_A
B	14	15	15	14	T_B
C	8	16	7	11	T_C
Total	B_1	B_2	B_3	B_4	$G=144$

$C.F = G^2/rk = (144)^2/12 = 1728.$

TSS - 130

$$SST = \frac{T_1^2 + T_2^2 + T_3^2}{4} = 38 \quad ; \quad SSB = \frac{B_1^2 + B_2^2 + B_3^2 + B_4^2}{3} = 62$$

ANOVA table for RBD.

Source of Variation	d.f	SS	MSS	F ratio: Fcal	F ratio: Ftab (at 5% loss)
Treatment	2	38	19	$3.8 = \hat{F}_T$	$F_{(2,6)} = 5.14$
Block	3	62	20.67	$4.13 = F_B$	$F_{(3,6)} = 4.76$
Error	6	30	5		
Total	$12-1=11$	130			

2) Carryout the analysis of the following design.

Blocks

I	II	III	IV
A 8	C 10	A 6	B 10
C 12	B 8	B 9	A. 8
B 10	A 8	C 10	C 9.

	I	II	III	IV	
A	8	8	6	8	30
B	10	8	9	10	37
C	12	10	10	9	41
	30	26	25	27	108 = G

$$C.F = \frac{(108)^2}{12} = \underline{\underline{972}}$$

$$RSS = 998$$

$$TSS = 998 - 972 = \underline{\underline{26}}$$

$$SST = \sum_{i=1}^{k} \frac{T_i^2}{k_i} - C.F = 15.5$$

$$SSB = 4.67 \; ; \quad SSE = 5.83$$

$$SSB = \sum_j \frac{B_j^2}{k} - C.F$$

$$SSE = TSS - SST - SSB.$$

ANOVA table

Source of variation	d.f.	SS	MSS	F values: F_{cal}	F values: F_{tab}
Treatments	3-1=2	15.5	$7.7 = \frac{15.5}{2}$	$7.94 = F_T$	$F_{(2,6)}$ at 5% = 5.14
Blocks	4-1=3	4.67	$1.56 = \frac{4.67}{3}$	$1.61 = F_B$.	$F_{(3,6)}$ at 5% = 4.76.
Error	6	5.83	$0.97 = \frac{5.831}{6}$		

Total	$12-1=11$ $rk-1=11$	26

$H_{0T} = \alpha_A = \alpha_B = \alpha_C = 0.$

$H_{0B} = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.$ $\boxed{\alpha = 0.05}$

Alternative hypothesis. Not all α's & β's are equal.

Decision making : $\alpha = 5\% = 0.05$

Here $F_T = 7.94 > F_{(2,6),0.05}$

Accept H_{0B} at 5% | Reject H_{0T} at 5% LOS.

Significant

(Critical difference).

$$C.D = \sqrt{\frac{2.MSSE}{r}} \times t_{\alpha/2} \text{ at error d.f.}$$

one sided tail in hypothesis testing.

$$= \sqrt{\frac{2 \times 0.97}{4}} \times t_{(0.025, 6)}$$

$$C.D = 0.69 \times 2.447 = 1.688$$

Pair of treatments	Abs dif. of means	C.D	Inference
(A, B)	$\lvert 7.5 - 9.25 \rvert = 1.76$	1.68	significant 1.76
(A, C)	$\lvert 9.25 - 10.25 \rvert = 1$	1.68	insignificant
(B, C)	$\lvert 7.5 - 10.25 \rvert = 2.75$	1.68	2.75

$\bar{A} = 30/4 = 7.5$, $\bar{B} = 37/4 = 9.25$, $\bar{C} = \frac{41}{4} = 10.25.$

Missing Plot Technique in RBD.

Let the observation y_{ij} in the jth block G_j recieving the ith treatment is missing.

	Treatments 1	2	- - - -	i	- - -	k	Total
1	y_{11}	y_{21}		y_{i1}		y_{k1}	B_1
2	y_{12}	y_{22}		y_{i2}		y_{k2}	B_2
⋮							
j	y_{1j}	y_{2j}		missing $\boxed{y_{ij}=x}$		y_{kj}	$B_j'+x=B_j$
⋮							
r	y_{1r}	y_{2r}		y_{ir}		y_{kr}	B_r
	T_1	T_2		$T_i'+x=T_i$ - - -		T_k	

Block

B_j' = Sum of known observation in jth block

T_i' = Sum of known observation in the ith treatment.

Total SS = TSS = RSS − C.F

$$= \sum_i \sum_j y_{ij} - \text{C.F.}$$

$$TSS = x^2 + \text{constant wrto } x - \text{C.F.}$$

$$SST = \frac{1}{r}\left[(T_i'+x)^2 + (\text{const wrt to } x)^2\right] - \text{C.F.}$$

$$SST = \frac{1}{r}\sum_i T_i^2 - \text{C.F}$$

$$TSS = RSS - C.F$$

$$= \sum_i \sum_j y_{ij}^2 - \frac{G^2}{n}$$

$$= \sum_i \sum_j (y_{ij}'^2 - x^2) - \frac{(G' + x)^2}{n}$$

$= x^2 +$ Constant terms wrt to $x -$ C.F.

$$SST = \frac{1}{r}\left[(T_i' + x)^2 + \text{const wrt } x\right] - C.F.$$

$$SSB = \frac{1}{k}\left[(B_j' + x)^2 + \text{const wrt } x\right) - C.F$$

where $CF = \frac{(G' + x)^2}{n}$

$$SSE = TSS - SST - SSB.$$