Challenging Experiment 9 and 10

(Design of Experiments –CRD, RBD and LSD)

Completely Randomized Design (in Design of Experiments):

The conducting of an experiment by allotting treatments (factors), whose effects are to be experimented, to uniform/homogeneous experimental units by a simple random sampling design such that every unit can receive any treatment with equal chance, analyzing (splitting) total variation in the results into variation due to treatments and variation due to chance (error or residual) and then testing the significance or otherwise of treatments variation over error variation.

CRD or Allotment of Treatments in CRD: It can be explained easily with 4 treatments denoted by the letters A, B, C, D to be allotted to 16 experimental units of uniform quality.

Let A be repeated on 3 units; B on 6 units; C on 3 units; D on 4 units.

Random allotment (randomization) is done like this-----serialize the 16 units from 1 to 16 first; take 16 slips of paper of equal size and shape, write the letters A on 3 slips. B on 6 slips, C on 3 slips and D on 4 slips; fold/roll these slips well separately, put them in a box/bag, shuffle them well and take one slip after another; allot the first chosen letter (slip) to the first unit, 2nd chosen letter to 2nd unit and so on until all the letters (treatments) from the chosen slips are allotted to all the 16 units completely.

NOTE: *Randomization of letters can also be done by ' Random Number Tables' or by 'computerized randomization'.*

Then the resulting layout may be as shown below:

CRD- layout in square form.

B	*D*	*A*	*C*
A	*C*	*B*	*B*
D	*B*	*C*	*A*
B	*D*	*B*	*D*

CRD-layout may be in rectangular form also as shown below: Let us consider treatment A repeated 4 times, B repeated 2 times, C repeated 5 times and D repeated 4 times on 15 units.

CRD ANOVA TABLE:-

Source of variation	df	SS	Mean SS= MSS = $\frac{SS}{df}$	F- ratio
Treat-ments	r–1	$\sum_{i=1}^{r}\left(\frac{R_i^2}{n_i}\right)$–CF	$\left(\frac{Treatment\ SS}{r-1}\right)$	$\left(\frac{Treatt.MSS}{Error\ MSS}\right)\to F_{[(r-1),\ (\sum n_i - r)]}$
Error	$\sum_{i=1}^{r}(n_i)-r$	by subtraction	$\left\{\frac{Error\ SS}{\sum_{i=1}^{r}(n_i)-r}\right\}$	------
Total	$\sum_{i=1}^{r}(n_i)-1$	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij}^2-CF$	-----	

The F-ratio =$\left(\frac{\text{Treatt.MSS}}{\text{ErrorMSS}}\right) \to F_{\left(\overline{r-1},\ \overline{\sum n_i - r}\right)}$- distribution with (r–1) as the numerator degrees of freedom (df) and $[(\sum_{i=1}^{r} n_i) - r]$ as the denominator df.

In this table, Correction Factor, CF = $\left[\frac{(\text{Grand Total})^2}{\text{Total no.of observations}}\right]$.

Complete Randomised design (Method of ANOVA for one way-classificantion with equal number of Observations)

Problem: Suppose the following table represents the sales figures of the 3 new menu items in the 18 restaurants after a week of test marketing. At .05 level of significance, test whether the average sales volume for the 3 new menu items are all equal.

Item 1	Item2	Item3
22	52	16
42	33	24
44	8	19
52	47	18
45	43	34
37	32	39

(Enter the Above data in Excel Sheet)

R code:-

```
> df1=read.csv("C:\\Users\\aadmin\\Desktop\\CRD.csv")
> df1
  Item.1 Item2 Item3
1     22    52    16
2     42    33    24
3     44     8    19
4     52    47    18
5     45    43    34
6     37    32    39
> r = c(t(as.matrix(df1)))   #resopnse data
> r
 [1] 22 52 16 42 33 24 44  8 19 52 47 18 45 43 34 37 32 39
> f = c("Item1", "Item2", "Item3")    #treatment levels
> f
[1] "Item1" "Item2" "Item3"
> k = 3      # number of treatment levels
> n = 6      # observations per treatment
> tm = gl(k, 1, n*k, factor(f)) # matching treatments
> tm
 [1] Item1 Item2 Item3 Item1 Item2 Item3 Item1 Item2 Item3 Item1 Item2 Item3
[13] Item1 Item2 Item3 Item1 Item2 Item3
Levels: Item1 Item2 Item3
> crdfit = aov(r ~ tm)
> summary(crdfit)
            Df Sum Sq Mean Sq F value Pr(>F)
tm           2  745.4   372.7   2.541  0.112
Residuals   15 2200.2   146.7
```

Interpretation : *Since the p-value of 0.112 is greater than the .05 significance level, we do not reject the null hypothesis that the mean sales volume of the new menu items are all equal*.

Completele Randomised design (Method of ANOVA for one way-classificantion with un equal number of Observations)

Problem :The following Table shows the lives(in hours) of four batches of electric lamps

Batches	Life of Bulbs in Hrs							
1	1600	1610	1650	1680	1700	1720	1800	
2	1580	1640	1640	1700	1750			
3	1460	1550	1600	1620	1640	1660	1740	1820
4	1510	1520	1530	1570	1600	1680		

(Enter the Above data in Excel Sheet)

R code:-

```
>data=c(1600,1610,1650,1680,1700,1720,1800,1580,1640,1640,1700,1750,1460,1550,1600,1620,1640,1660,1740,1820,1510,1520,1530,1570,1600,1680)

>batchs=c("batch1","batch1","batch1","batch1","batch1","batch1","batch1","batch2","batch2","batch2","batch2","batch2","batch3","batch3","batch3","batch3","batch3","batch3","batch3","batch3","batch4","batch4","batch4","batch4","batch4","batch4")

> Anova1=aov(data~batchs)

> summary(Anova1)

          Df Sum Sq Mean Sq F value Pr(>F)
batchs     3  44361   14787   2.149 0.123
Residuals 22 151351    6880
```

Interpretation :

We may regard the four batches of electric lamps to be homogeneous

Randomized Block Design (in Design of Experiments): The conducting of an experiment on experimental units, which differ in quality with respect to one character and hence stratified with respect to such changing character into different within-homogeneous blocks/strata and then allotting treatments, whose effects are to be experimented, to homogenized units within each block by simple random sampling design independently without repetitions and thereafter splitting the total variation in the results into blocks variation, treatments variation, residual /error variation and lastly testing the significance of these variations over error variation.

In symbols,

allotting treatments to experimental units, differing in quality with respect to one character, by stratified random sampling, by which

Total variation = Blocks variation + Treatments variation + Error variation

and then testing if Blocks variation and Treatments variation are significant over Error variation, obtained from the results on such randomized block design(RBD) data .

A specimen of RBD, with four treatments A,B,C,D in 5 blocks taken in rows, say is given below:

with rows as blocks---→

Block-I	D	C	A	B
Block-II	B	A	C	D
Block-III	A	D	B	C
Block-IV	C	D	B	A
Block-V	B	A	D	C

Or

With columns as blocks---→

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

R.B.D ANOVA table:-

Source of Variation	**df**	**SS**	**Mean SS=** $\frac{SS}{df}$	**F-ratio (F-distribution)**
Rows (Blocks)	r – 1	$\frac{\sum_{i=1}^{r} R_i^2}{c}$ – CF	$\frac{\text{Row SS}}{\text{r} - 1}$	$\left(\frac{\text{Row MSS}}{\text{Error MSS}}\right) \rightarrow F_{[(r-1),\ (r-1)(c-1)]}$
Columns (Treatm-ents)	c – 1	$\frac{\sum_{j=1}^{c} C_j^2}{r}$ – CF	$\frac{\text{Column SS}}{\text{c} - 1}$	$\left(\frac{\text{Column MSS}}{\text{Error MSS}}\right) \rightarrow F_{[(c-1),\ (r-1)(c-1)]}$
Error	(r –1) (c –1)	(by subtra-ction)	$\frac{\text{Error SS}}{(r-1)(c-1)}$	-----
Total	rc – 1	$\sum_{i=1}^{r}\sum_{j=1}^{c} {x_{ij}}^2$ – CF	-----	

In the table, MSS = Mean SS, CF = Correction factor = $\left[\frac{(GT)^2}{\text{Total no.of observations, } rc}\right]$ and

df = degrees of freedom.

The F-ratio = $\left(\frac{\text{Row MSS}}{\text{Error MSS}}\right) \rightarrow F_{[(r-1),\ (r-1)(c-1)]}$ –distribution with (r – 1) as the numerator df and (r – 1)(c – 1) as the denominator (Error) df.

Similarly, the F-ratio = $\left(\frac{\text{Column MSS}}{\text{Error MSS}}\right) \rightarrow F_{[(c-1),\ (r-1)(c-1)]}$ –distribution with (c –1) as the numerator df and (r – 1)(c – 1) as the denominator (Error) df.

Two-way analysis of variance:

Problem: Suppose each row in the following table represents the sales figures of the 3 new menu in a restaurant after a week of test marketing. At .05 level of significance, test whether the average sales volume for the 3 new menu items are all equal.

Data :-

ITEM1	ITEM2	ITEM3
31	27	24
31	28	31
45	29	46
21	18	48
42	36	46
32	17	40

```
> df2=read.csv("C:\\Users\\aadmin\\Desktop\\RBD.csv")
> df2
  ITEM1 ITEM2 ITEM3
1    31    27    24
2    31    28    31
3    45    29    46
4    21    18    48
5    42    36    46
6    32    17    40
> r = c(t(as.matrix(df2))) # response data
> r
 [1] 31 27 24 31 28 31 45 29 46 21 18 48 42 36 46 32 17 40
> f = c("Item1", "Item2", "Item3") # treatment levels
> k = 3 # number of treatment levels
> n = 6 # number of control blocks
> tm = gl(k, 1, n*k, factor(f)) # matching treatment
> tm
 [1] Item1 Item2 Item3 Item1 Item2 Item3 Item1 Item2 Item3 Item1 Item2 Item3
[13] Item1 Item2 Item3 Item1 Item2 Item3
Levels: Item1 Item2 Item3
> blk = gl(n, k, k*n) # blocking factor
> blk
 [1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6
Levels: 1 2 3 4 5 6
> rbdfit = aov(r ~ tm + blk)
> summary(rbdfit) # Print out the ANOVA table with the summary function.
            Df Sum Sq Mean Sq F value Pr(>F)
tm           2  538.8  269.39   4.959 0.0319 *
blk          5  559.8  111.96   2.061 0.1547
Residuals   10  543.2   54.32
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation :-

Since the p-value of 0.032 is less than the .05 significance level, we reject the null hypothesis that the mean sales volume of the new menu items are not all equal.

Problem 2 : The data recorded for yield in a randomized block design experiment involving six treatments in four randomized blocks are given below. Analyse and interpret the data

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

Solution : Arrange this data in order

		Treatments and yield					
		1	**2**	**3**	**4**	**5**	**6**
Blocks	1	24.7	20.6	27.7	16.2	16.2	24.9
	2	27.3	28.8	22.7	15.0	17.0	22.5
	3	38.5	39.5	36.8	19.6	15.4	26.3
	4	28.5	31.0	34.9	14.1	17.7	22.6

R code:

```
> data=c(24.7,20.6,27.7,16.2,16.2,24.9,27.3,28.8,22.7,15.0,17.0,22.5,38.5,39.5,36.8,19.6,15.4,26.3,28.5,31.0,34.9,14.1,17.7,22.6)
> blocks=gl(4,6)
> treatments=gl(6,1,24)
> rbdfit=aov(data~blocks+treatments)
> rbdfit
Call:
   aov(formula = data ~ blocks + treatments)

Terms:
                  blocks treatments Residuals
Sum of Squares  219.4279   901.1921  229.6396
Deg. of Freedom        3          5        15

Residual standard error: 3.912711
Estimated effects may be unbalanced
> summary.aov(rbdfit)
            Df Sum Sq Mean Sq F value   Pr(>F)
blocks       3  219.4   73.14   4.778  0.0157 *
treatments   5  901.2  180.24  11.773 9.28e-05 ***
Residuals   15  229.6   15.31
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation :-

Here P< 0.05.Then Blocks are not homogenous .Treatments effects are not alike .

Latin Square Design (in Design of Experiments): The conducting of an experiment on r^2 experimental units, which differ in quality with respect to two characters and hence stratified, once with respect to one changing character, into different within-homogeneous rows and again, with respect to second changing character, into different within-homogeneous columns arranged in r^2-square form and then allotting 'r' treatments (whose effects are to be experimented) denoted by the Latin letters A, B, C, D, . . . to such within-homogenized units in each row separately without repetitions and also in each column separately without repetitions, splitting total variation in results into row-factor variation, column-factor variation, treatments variation, residual (error) variation and lastly testing the significance of these variations over error variation (variance).

In symbols, an experiment in which

Total variation =

Row-factor variation + Column-factor variation + Treatments variation+ Error variation

and then these variations are tested for their significance over error variation obtained from the results on applying r-treatments on r^2-units, without repetitions in any row or column, arranged in a square, which differ in their quality with respect to two characters, whose specimen with the treatments denoted in the Latin letters A, B, C & D, is given below:

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Anova table for L.S.D:

Source of vari-ation	df	SS	Mean SS $= \left(\frac{\text{SS}}{\text{df}}\right)$	F-ratio (F-distribution)
Row-factor	r–1	$\sum_{i=1}^{r}\left(\frac{R_i^2}{r}\right)$ – CF	$\left(\frac{\text{Row SS}}{\text{r-1}}\right)$	$\left(\frac{\text{Row M SS}}{\text{Error MSS}}\right) \rightarrow$ $\text{F}_{[\text{r-1},\ (\text{r-1})(\text{r-2})\}}$
Column factor	r–1	$\sum_{j=1}^{r}\left(\frac{C_j^2}{r}\right)$ – CF	$\left(\frac{\text{Column SS}}{\text{r-1}}\right)$	$\left(\frac{\text{Column M SS}}{\text{Error MSS}}\right) \rightarrow$ $\text{F}_{[\text{r-1},\ (\text{r-1})(\text{r-2})\}}$
Treat-ments	r–1	$\sum_{k=1}^{r}\left(\frac{T_k^2}{r}\right)$ – CF	$\left(\frac{\text{Treatments SS}}{\text{r-1}}\right)$	$\left(\frac{\text{Treatments M SS}}{\text{Error MSS}}\right) \rightarrow$ $\text{F}_{[\text{r-1},\ (\text{r-1})(\text{r-2})\}}$
Error	(r–1)(r–2)	By subtraction	$\left(\frac{\text{Error SS}}{(\text{r-1})(\text{r-2})}\right)$	-----
Total	r^2–1	$\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r} {x_{ijk}}^2$ – CF	-----	

In the above table, df = degrees of freedom ,SS = sum of squares, MSS = Mean SS = $\left(\frac{SS}{df}\right)$, CF = Correction factor = $\frac{(\text{Grand Total})^2}{\text{Total no. of observations}}$.

The F-ratio = $\left(\frac{\text{Row/Column/Treatment MSS}}{\text{Error MSS}}\right) \rightarrow F_{[r-1,\ (r-1)(r-2)\}}$ -distribution with (r–1) as the numerator df and (r–1)(r–2) as the denominator (Error) df. In the above ANOVA table, row SS +column SS + treatments SS + Error SS = total SS, but their MSS do not add together to give total MSS.

Latin Square Design:-

Suppose we want to analyze the productivity of 5 kinds of fertilizers, 5 kinds of tillage(land under cultivation) and 5 kinds of seeds. The data is organized in a LSD as follows

	Treat A	***Treat B***	***Treat C***	***Treat D***	***Treat E***
Fertilizer 1	***"A42"***	***"C47"***	***"B55"***	***"D51"***	***"E44"***
Fertilizer 2	***"E45"***	***"B54"***	***"C52"***	***"A44"***	***"D50"***
Fertilizer 3	***"C41"***	***"A46"***	***"D57"***	***"E47"***	***"B48"***
Fertilizer 4	***"B56"***	***"D52"***	***"E49"***	***"C50"***	***"A43"***
Fertilizer 5	***"D47"***	***"E49"***	***"A45"***	***"B54"***	***"C46"***

R code:

```
> fertil <- c(rep("fertil1",1), rep("fertil2",1), rep("fertil3",1), rep("fertil4",1),
rep("fertil5",1))
> treat <- c(rep("treatA",5), rep("treatB",5), rep("treatC",5), rep("treatD",5),
rep("treatE",5))
> seed <- c("A","E","C","B","D", "C","B","A","D","E", "B","C","D","E","A",
"D","A","E","C","B", "E","D","B","A","C")
> freq <- c(42,45,41,56,47, 47,54,46,52,49, 55,52,57,49,45, 51,44,47,50,54,
44,50,48,43,46)
> mydata <- data.frame(treat, fertil, seed, freq)
> myfit <- lm(freq ~ fertil+treat+seed, mydata)
> anova(myfit)
```

```
Analysis of Variance Table

Response: freq
          Df Sum Sq Mean Sq F value    Pr(>F)
fertil     4  17.76   4.440  0.7967 0.549839
treat      4 109.36  27.340  4.9055 0.014105 *
seed       4 286.16  71.540 12.8361 0.000271 ***
Residuals 12  66.88   5.573
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation:

– The difference between group considering the fertilizer is not significant (p-value > 0.1);
– The difference between group considering the tillage is quite significant (p-value < 0.05);
– The difference between group considering the seed is very significant (p-value < 0.001);

Practice Questions and Challenging Experiments:-

Completely randomised design:-

1. A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee

Program 1	Program 2	Program 3	Program 4
9	10	12	9
12	6	14	8
14	9	11	11
11	9	13	7
13	10	11	8

Calculate and interpret the above one way ANOVA table.

2. In a factory producing edible oil and marketing its product in 15 kg tins, uses five filling machines. Random samples of the packed tins were taken for each machine A,B,C,D and E were presented below

A	B	C	D	E
14.85	14.28	14.16	15.25	14.60
15.00	14.42	14.15	15.30	14.84
15.25		14.19	15.10	14.82
15.10		14.50	15.35	14.74
14.80			15.00	

Analysis of data to test the Equality of efficiency of machines .

Randomised Block Design (R.B.D)

A factory manager wanted to compare the efficiency of four factory workers with respect to cotton spinning. He had four machines .Four workers A,B,C and D were allotted a machine as experimental units. Cotton thread yield(in kg) for each worker was recorded as shown in the layout displayed below. All machines were of the same make

Blocks	Machines				
I	C(22)	A(14)	B(12)	A(23)	
II	A(16)	B(18)	C(20)	D(25)	
III	A(15)	C(23)	D(28)	B(14)	
IV	A(17)	D(21)	C(19)	B(11)	
V	B(18)	C(20)	A(13)	D(24)	
VI	D(18)	C(21)	A(10)	B(17)	

Latin Square Design (L.S.D)

Analyse and interpret the following statistics concerning output of wheat per field obtained as a result of experiment conducted to test four varieties of wheat Viz., A,B,C and D under a Latin Square Design

C (25)	**B (23)**	**A(20)**	**D (20)**
A (19)	**D (19)**	**C (21)**	**B (18)**
B (19)	**A (14)**	**D (17)**	**C (20)**
D (17)	**C (20)**	**B (21)**	**A (15)**