2. What do you mean by Latin Square Design? Write down its merit and demerit. Set up the analysis of variance for the following of design.

A (10)	B (15)	C(20)
B (25)	C (10)	A (15)
C (25)	A (20)	B (15)

Ans: When the experimental material is not homogeneous the LSD is better than RBD. In RBD local control is used according to one way grouping i.e. according to blocks but in LSD local control is used according to two way grouping i.e. rows and columns. Hence it is used when two sources of errors are to be controlled simultaneously. In this design number of treatments are equal to the number of replication and the treatments are allocated in such a way that each of the treatment occurs once and only once in each row and column. In this design Latin alphabet are used to denote the treatments, and shape is square due to equal number of treatments and replication so called Latin square design. It is based upon the all principles of design namely replication, randomization and local control.

Let us consider m treatments with m replication each so that there are $N = m^2$experimental unit.

Let us divide the experimental material into m^2 experimental units arranged in square so that each row as well as column contains m units. In this design none of treatments are replicated along row wise or column wise. In this case we study the variation between treatments, the variation between rows and variation between columns. It has only m^2 experimental unit but studies variation of three factors i.e. rows, columns and treatments. Hence it is the case of incomplete three way ANOVA. For complete three way ANOVA we need m^3 experimental unit.

Let us consider t = 4(A, B, C, D) then 4 × 4 LSD is as shown below.

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

Merits of LSD

i) Due to the use of two way grouping of controls more variation than CRD and RBD.

ii) It is incomplete three way layout. It's advantage over complete three way layout is that instead of m^3 experimental units only m^2 units are needed.

iii) The statistical analysis remains simple if some observations are missing.

Demerits of LSD

i) The assumption of factors are independent is not always true.

ii) It is suitable for treatments 5 to 10.

iii) It is not easy in the field layout.

Now,

Problem to test

H_{0R} : Rows are insignificant
H_{1R} : Rows are significant
H_{0C} : Columns are insignificant
H_{1C} : Columns are significant
H_{0T} : Treatments are insignificant
H_{1T} : Treatments are significant

				$T_{i..}$	$T_{i..}^2$
	A 10	B 15	C 20	45	2025
	B 25	C 10	A 15	50	2500
	C 25	A 20	B 15	60	3600
$T_{.j.}$	60	45	50	G =155	$\sum T_{i..}^2$ =8125
$T_{.j.}^2$	3600	2025	2500	$\sum T_{.j.}^2$ = 8125	

$T_{..A}$ = 10+20+15 = 45
$T_{..B}$ = 25+15+15 = 55
$T_{..C}$ = 25+10+20 = 55
$\sum T_{..k}^2 = 45^2 + 55^2 + 55^2 = 8075$
k = A , B , C
m = 3
$N = m^2 = 3^2 = 9$
$CF = \frac{G^2}{N} = \frac{155^2}{9} = 2669.444$
$\sum y_{ijk}^2 = 10^2 + 15^2 + 20^2 + 25^2 + 10^2 + 15^2 + 25^2 + 20^2 + 15^2 = 2925$

$TSS = \sum_{(i,j,k)} y_{ijk}^2 - C.F. = 2925 - 2669.444 = 255.556$

$SSR = \frac{\Sigma_i T_{i..}^2}{m} - C.F. = \frac{1}{3} \times 8125 - 2669.444 = 38.889$

$SSC = \frac{\Sigma_j T_{.j.}^2}{m} - C.F. = \frac{1}{3} \times 8125 - 2669.444 = 38.889$

$SST = \Sigma \frac{T_{..k}^2}{m} - C.F. = \frac{1}{3} \times 8075 - 2669.444 = 22.222$

SSE = TSS – SSR – SSC – SST
= 255.556 – 38.889 – 38.889 – 22.222 = 155.556

ANOVA table

S.V.	d.f.	S.S.	M.S.	F_{Cal}	F_{Tab}
Row	2	38.889	19.444	0.249	$F_{0.05(2,2)} = 19$
Column	2	38.889	19.444	0.249	$F_{0.05(2,2)} = 19$
Treatment	2	22.222	11.111	0.142	$F_{0.05(2,2)} = 19$
Error	2	155.556	77.778		
Total	8	255.556			

Decision

$F_R = 0.249 <> F_{0.05(2,2)} = 19$, Accept H_{0R} at 5% level of significance.
$F_C = 0.249 <> F_{0.05(2,2)} = 19$, Accept H_{0C} at 5% level of significance.
$F_T = 0.142 <> F_{0.05(2,2)} = 19$, Accept H_{0T} at 5% level of significance.

Conclusion

Rows are insignificant.

Columns are insignificant.

Treatments are insignificant.

3. **What do you mean by hypothesis? Describe null and alternative hypothesis. A company claims that its light bulbs are superior to those of the competitor on the basis of study which showed that a sample of 40 of its bulbs had an average life time 628 hours of continuous use with a standard deviation of 27 hours. While sample of 30 bulbs made by the competitor had and average life time 619 hours of continuous use with a standard deviation of 25 hours. Test at 5% level of significance, whether this claim is justified.**

Ans: A hypothesis is a tentative theory or supposition provisionally adopted to explain certain facts and to guide in the investigation of others.

A statistical hypothesis which is tentative statement or supposition about the estimated value of one or more parameter of the population is called parametric hypothesis. A statistical hypothesis about attributes is called non-parametric hypothesis.

If a hypothesis completely determines the population, it is called a simple hypothesis, otherwise composite hypothesis.

In testing of hypothesis a statistic is computed from a sample drawn from the parent population and on the basis of the statistic it is observed whether the sample so drawn has come from the population with certain specified characteristic.

Null hypothesis

The supposition about the population parameter is called null hypothesis. It is set up for testing a statistical hypothesis only to decide whether to accept or reject the null hypothesis. According to R.A. Fisher, null hypothesis is the hypothesis which is tested for possible rejection under the assumption that is true.

It is the hypothesis of no difference between sample statistic and parameter. It is hypothesis of no difference between parameters.

Null hypothesis is denoted by H_0. It is set up as $H_0: \mu = \mu_0$

Suppose we want to test the average score of students in B.Sc. entrance exam is 55 then to start testing the hypothesis we assume the average score is 55. There is no difference between sample average and population average. Then the null hypothesis is $H_0: \mu = 55$

Alternative Hypothesis

A hypothesis which is complementary to the null hypothesis is called an alternative hypothesis.

Any hypothesis which is not null is also called alternative hypothesis. It is hypothesis of difference between sample statistic and parameter. It is hypothesis of difference between parameters.

Alternative hypothesis is denoted by H_1.

Alternative hypothesis are

(i) two tailed

(ii) one tailed right

(iii) one tailed left

Alternative hypothesis is set up as $H_1: \mu \neq \mu_0$ for two tailed or $H_1: \mu > \mu_0$ for one tailed right or $H_1: \mu < \mu_0$ for one tail left.

Let μ_1 *and* μ_2 be mean life of bulbs of company and it's competitor respectively

Sample number of bulb of company(n_1) = 40

Sample mean life of bulb of company ($\bar{X}_1$) = 628,

Sample Sd of life of bulb of company(s_1) = 27

Sample number of bulb of competitor (n_2) = 30,

Sample mean life of bulb of competitor ($\bar{X}_2$) = 619

Sample Sd of life of bulb of competitor (s_2) =25

Let μ_1 = Population mean wage of workers from Pokhara and μ_2 = Population mean wage of workers from Kathmandu.

Problem to test

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

Test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{628-619}{\sqrt{\frac{27^2}{40}+\frac{25^2}{30}}} = \frac{9}{\sqrt{39.058}} = 1.44$$

Z = 1.44

Critical value

At α = 0.05 be the level of significance then the critical value for one tailed test is

$Z_{tabulated} = Z_\alpha = 1.645$

Decision

$Z = 1.44 < Z_{tabulated} = 1.645$, accept H_0 at 0.05 level of significance

Conclusion

The claims of company that its light bulbs are superior to those of the competitor is not correct.

Group 'B'

Attempt any Eight questions: **(8 × 5 = 40)**

4. Suppose we are given following information with n = 7, multiple regression mode is $\hat{Y} = 8.15 + 0.56X_1 + 0.54X_2$

Here, Total sum of square = 1493,

Sum of square due to error = 91

Find i) R^2 and interpret it. ii) Test the overall significance of model.

Ans:

TSS = 1493

SSE = 91

SSR = TSS – SSE = 1493 – 91 = 1402

MSR = SSR/k = 1402/2 = 701

MSE = SSE/n-k-1 = 91/7-2-1 = 22.75

$R^2 = SSR/TSS = 1402/1493 = 0.939 = 93.9\%$

It means 93.9% variation in y is explained by x_1 *and* x_2

To test overall significance of regression model

Let β_1 *and* β_2 be population regression coefficient of Y on X_1 keeping X_2 constant and population regression coefficient of Y on X_2 keeping X_1 constant

Problem to test

$H_0 : \beta_1 = \beta_2 = 0$

H_1 : At least one β_i is different from zero, i = 1, 2

Test statistic

$$F = \frac{MSR}{MSE} = 701/22.75 = 30.81$$

Critical value

At $\alpha = 0.05$ level of significance, critical value is $F_{\alpha(k, n-k-1)} = 6.944$

Decision

$F = 30.81 > F_{tabulated} = 6.944$, reject H_0 at 5% level of significance.

Conclusion

There is linear relationship of dependent variable y with at least one of the independent variable x's

5. The following data related to the number of children classified according to the type of feed and the nature of teeth.

Type of feed	Nature of Teeth	
	Normal	Defective
Breast	18	12
Bottle	2	13

Do the information provide sufficient evidence to conclude that type of feeding and nature of teeth are dependent? Use chi square test at 5% level of significance

Ans: **Problem to test**

H_0 : Type of feeding and nature of teeth are independent

Against H_1 : Type of feeding and nature of teeth are dependent

Type of feed	Nature of teeth		
	Normal	Defective	Total
Breast	a=18	b=12	a+b=30
Bottle	c=2	d=13	c+d=15
Total	a+c=20	b+d=25	N=45

Test statistic

$$\chi^2 = \frac{N(|ad - bc| - \frac{N}{2})^2}{(a + c)\,(b + d)\,(a + b)\,(c + d)}$$

$$= \frac{45(|18\times13-12\times2|-45/2)^2}{20\times25\times30\times15} = 7.031$$

Critical value

At $\alpha = 0.05$ level of significance critical value is $\chi^2_{\alpha(1)} = 3.84$

Decision

$\chi^2 = 7.031 > \chi^2_{0.01(1)} = 3.84$

Reject H_0 at 0.05 level of significance

Conclusion

Type of feeding and nature of teeth are dependent

6. **Determine the minimum sample size required so that the sample estimate lies within 10% of the true value with 95% level of confidence when coefficient of variation is 60%.**

Ans: Here, C.V. = 60% = 0.6

$P(|\bar{x} - \mu| \leq 0.1\mu) = 0.95$ ......(i)

Confidence level (1-α) = 95% = 0.95 then α = 0.05

Now, $P(|\bar{x} - \mu| \leq \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}) = 1 - \alpha$

⇨ $P(|\bar{x} - \mu| \leq 1.96 \times \frac{\sigma}{\sqrt{n}}) = 0.95$ ......(ii)

From equation (i) and (ii)

$0.1\mu = 1.96 \times \frac{\sigma}{\sqrt{n}}$

⇨ $\sqrt{n} = \frac{1.96}{0.1} \times \frac{\sigma}{\mu}$

⇨ $n = (1.96/0.1 \times \sigma/\mu)^2$

⇨ $n = 384.16 \times CV^2$

⇨ $n = 384.16 \times (0.6)^2$

⇨ $n = 138.29 \approx 138$

Hence required sample size is 138.

7. **A manufacture of computer paper has a production process that operates continuously throughout an entire production shift. The paper is expected to have an average length of 11inches and standard deviation is known to be 0.01 inch. Suppose random sample of 100 sheets is selected and the average paper length is found to be 10.68 inches. Set up 95% and 90% confidence interval estimate of the population average paper length.**

Ans: Average length of paper (μ) = 11

Standard deviation of paper (σ) = 0.01

Sample of paper sheet (n) = 100

Sample average length of paper($\bar{x}$) = 10.68

Level of confidence (100 - Θ)% = 95%

Level of significance (Θ) = 5%

Level of confidence (100 - Θ)% = 90%

Level of significance(Θ) = 10%

Level of significance

$\alpha = 5\% = 0.05$

Critical value

From Kruskal Wallis table critical value is $p = 0.009$

Decision

$P = 0.009 < \alpha = 0.05$, reject H_0 at 0.05 level of significance.

Conclusion

There is at least one significant difference between catalyst.

9. **Consider the partially completed ANOVA table below. Complete the ANOVA table and answer the following:**

Source of Variation	Sum of Square	Degree of freedom	Mean sum of square	F value
Column	72	?	?	2
Rows	?	?	36	?
Treatments	180	3	?	?
Error	?	6	12	
Total	?	?		

i. What design was employed?

ii. How many treatments were compared?

Ans: Design LSD was employed

Here df for treatment = m-1 = 3 , hence m = 4

Hence 4 treatments were compared

df for columns = m-1 = 3

df for rows = m-1 = 3

Total number of observations = $m^2 = 16$

df for total = $m^2 - 1 = 16\text{-}1 = 15$

MSC = SSC/m-1 = 72/3 = 24

MST = SST/m-1 = 180/3 = 60

SSR = (m-1) MSR = 3× 36 = 108

SSE = (m-1)(m-2) MSE = 6× 12 = 72

TSS = SSC + SSR + SST + SSE = 72+108+180+72= 432

F_C = MSC/MSE = 24/12 = 2

F_R = MSR/MSE = 36/12 = 3

F_T = MST/MSE = 60/12 = 5

Source of Variation	Sum of Square	Degree of freedom	Mean sum of square	F value
Column	72	3	24	2
Rows	108	3	36	3
Treatments	180	3	60	5
Error	72	6	12	
Total	432	15		

10. **Defined main component of queuing system.**

Ans: Components of queuing system are as explained below

Arrival

Job arrives to the queuing system at random times. A counting process A(t) tells the number of arrivals that occurred by time t . In stationary queuing system arrivals occur at arrival rate