Assignment-3

Date: 2081
Page:

Q1) What do you mean by Design of Experiment ? Physicians depend the laboratory test results when managing medical problems such as diabetes or epilepsy. In a uniformity test glucose tolerance, 3 different laboratories were each $n_i = 5$ identical blood samples from a person who had drunk 50 mg of glucose in dissolved water. The laboratory results (mg/d.l) are listed here :

lab1	lab2	lab3
12.1	9.3	10.0
11.7	11.1	10.5
10.9	10.7	10.1
10.2	10.9	11.0
10.6	9.01	10.4

Do the data indicate a difference in the average readings for 3 laboratories ? Use 0.05 level of significance.

→ Design of Experiment

↳ The objective of experiment is to maximize precision (information) and minimize error.

Design of experiment refers to systematic approach and structure employed in scientific experiments to investigate cause and effect relationships between variables. It involves careful manipulation of an independent variable while controlling and measuring other variables to assess their impact on dependent variables.

→ The numerical portion of the problem can be solved using one way ANOVA as follows :-

Date :
Page:

Let, lab1 = A, lab2 = B, lab3 = C.

Then,

						T_i
A	12.1	11.7	10.9	10.2	10.6	55.5
B	9.3	11.1	10.7	10.9	9.01	51.01
C	10.0	10.5	10.1	11.0	10.4	52

$G = \sum_i T_i = 158.51$

Here,

Problem to test

$H_0T: \mu_A = \mu_B = \mu_C$

H_1T: At least μ_i is different $[i = A, B, C]$.

Statistical Analysis

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

Where,

μ and T_i are parameters and are estimated using principle of least square by minimizing error sum of square.

Now,

From above table

No. of treatments $(t) = 3$

Replication $(r) = 5$

i.e. $N = rt = 5 \times 3 = 15$

Then,

Total sum of square (TSS) $= \sum_i \sum_j y_{ij}^2 - \frac{G^2}{N}$

Where,

$$\sum_i \sum_j y_{ij}^2 = 12.1^2 + 11.7^2 + 10.9^2 + 10.2^2 + 10.6^2 + 9.3^2 + 11.1^2 + 10.7^2 + 10.9^2 + 9.01^2 + 10.0^2 + 10.5^2 + 10.1^2 + 11.0^2 + 10.4^2$$

$= 1684.1101$

i.e. $TSS = 1684.1101 - \frac{(158.51)^2}{15} = 9.0821$

Similarly,

Sum of Square due to treatment (SST) $= \sum_i \frac{T_i^2}{r} - \frac{G^2}{N}$

Where, $\sum_i T_i^2 = 55.5^2 + 51.01^2 + 52^2 = 8386.2701$

i.e. $SST = \frac{8386.2701}{5} - \frac{(158.51)^2}{15}$

$= 2.2260$.

Then,

Sum of Square due to error (SSE) $= TSS - SST$

$SSE = 9.0821 - 2.2260$

$= 6.8561$

ANOVA Table

SV	df	SS	MS	Fcal	FTab
Treatment	2	2.2260	1.113	$F_T = 1.9480$	$F_{0.05}(2,12) =$ 3.885
Error	12	6.8561	0.5713		
Total	14	9.0821			

Decision

$F_T = 1.9480 < F_{0.05}(2,12) = 3.885$.

Accept H_{0T} at 0.05 level of significance

Conclusion

There is no difference in average reading for 3 laboratories.

CamScanner

Date :
Page:

Q2) State mathematically model for statistical analysis of mxm LSD for one observation per experimental unit. Also prepare dummy ANOVA table for this.

⇒ Latin Square Design (LSD)

It is used in heterogeneous material and uses all 3 principles of design namely replication, ramdomization and local Control. Treatments are not repeated along row and Column. It has square shape and latin letters are used to represent treatments, hence called latin Square Design.

* Mathematical Model for LSD

$y_{ijk} = \mu + r_i + c_j + T_k + e_{ijk}$

Where,

$i = 1, 2, 3, ---, m; \; j = 1, 2, 3 --- m; \; k = 1, 2, 3, ---, m$

y_{ijk} = Observation due to i^{th} row, j^{th} column & k^{th} treatment

μ = General Mean

c_j = Effect due to j^{th} Column

r_i = Effect due to i^{th} row

T_k = Effect due to k^{th} treatment

e_{ijk} = Error due to i^{th} row, j^{th} column & k^{th} treatment.

* Statistical Analysis of LSD

In the mathematical model of LSD, μ, r_i, c_j and T_k are parameters which can be estimated using principle of least square of minimizing error sum of square.

Date: ....
Page: ....

* Dummy ANOVA TABLE For m×m LSD

SV	df	SS	MS	Fcal	FTab
Row	$m-1$	SSR	$MSR = \frac{SSR}{m-1}$	$F_R = \frac{MSR}{MSE}$	$F_\alpha\{(m-1),(m-1)(m-2)\}$
Column	$m-1$	SSC	$MSC = \frac{SSC}{m-1}$	$F_C = \frac{MSC}{MSE}$	"
Treatment	$m-1$	SST	$MST = \frac{SST}{m-1}$	$F_T = \frac{MST}{MSE}$	"
Error	$(m-1)(m-2)$	SSE	$MSE = \frac{SSE}{(m-1)(m-2)}$		
Total	m^2-1	TSS			

Q3) What do you mean by LSD? Give mathematical model along with meanings. Following information represe[nts] yield on using different types of treatments.

A 12	C 16	B 11	D 20
B 21	D 19	C 15	A 13
C 16	A 13	D 16	B 19
D 10	B 14	A 19	C 16

Carry out analysis of the design

→ It is used in heterogeneous material and uses all 3 principles of design namely replication, randomization and local control. Treatments are not replicated along row as well as column. It has square shape and latin letters are used to represent treatments, hence called Latin Square Design.

Date : ....
Page: ....

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 insignificant

Here,

We are given,

				$T_{i..}$
A 12	C 16	B 11	D 20	59
B 21	D 19	C 15	A 13	68
C 16	A 13	D 16	B 19	64
D 10	B 14	A 19	C 16	59
$T_{.j.}$ 59	62	61	68	G = 250

And,

$i = A, B, C, D$; $j = A, B, C, D$.

$T_{..k}$ can be calculated as $[k = A, B, C, D]$

$T_{..A} = 12 + 13 + 13 + 19 = 57$

$T_{..B} = 11 + 21 + 19 + 14 = 65$

$T_{..C} = 16 + 15 + 16 + 16 = 63$

$T_{..D} = 20 + 19 + 16 + 10 = 65$; $T_{..k} = 250$

And,

$N = m^2 = 4^2 = 16$

$G = \sum_i T_{i..} = \sum_j T_{.j.} = \sum_k T_{..k} = 250.$

Date :
Page:

Then, $\sum_{i,j,k} y_{ijk}^2 = 12^2+16^2+11^2+20^2+19^2+21^2+15^2+13^2+16^2+13^2+16^2+19^2+10^2+14^2+19^2+16^2$

$= 4072$

So,

$$TSS = \sum_{i,j,k} y_{ijk}^2 - \frac{G^2}{N} = 4072 - \frac{250^2}{16} = 165.75$$

$$SSR = \sum_i \frac{T_{i..}^2}{m} - \frac{G^2}{N} = \frac{59^2+68^2+64^2+59^2}{4} - \frac{250^2}{16}$$

$$= 14.25$$

$$SSC = \sum_j \frac{T_{.j.}^2}{m} - \frac{G^2}{N} = \frac{59^2+62^2+61^2+68^2}{4} - \frac{250^2}{16}$$

$$= 11.25$$

$$SST = \sum_k \frac{T_{..k}^2}{m} - \frac{G^2}{N} = \frac{57^2+65^2+63^2+65^2}{4} - \frac{250^2}{16}$$

$$= 10.75$$

$$SSE = TSS - SSR - SSC - SST$$
$$= 165.75 - 14.25 - 11.25 - 10.75$$
$$= 129.5$$

ANOVA Table

let, $\alpha = 0.05$ be level of significance

SV	df	SS	MS	Fcal	Ftab
Row	3	14.25	4.75	$F_R = 0.2201$	$F_{0.05}(3,6) = 4.757$
Column	3	11.25	3.75	$F_C = 0.1737$	"
Treatment	3	10.75	3.5833	$F_T = 0.1660$	"
Error	6	129.5	21.5833		
Total	15	165.75			

Date :
Page:

Decision

$F_R = 0.2201 < F_{0.05}(3,6) = 4.757$
Accept H_0R at $\alpha = 0.05$

$F_C = 0.1737 < F_{0.05}(3,6) = 4.757$
Accept H_0C at $\alpha = 0.05$

$F_T = 0.1660 < F_{0.05}(3,6) = 4.757$
Accept H_0T at $\alpha = 0.05$

Conclusion

Rows are insignificant at $\alpha = 0.05$
Columns are insignificant
Treatments are insignificant.

Q4) Give layout of Completely Randomized Design. Write down mathematical model and ANOVA Table of the design.

→ **Completely Randomized Design**

It is used for homogeneous material and uses only 2 principles of design namely replication and randomization. In this case treatments are replicated row wise as well as column wise. It is case of one way ANOVA.

- layout, Consider

Treatments = 4 (A, B, C, D)
Replication = 4

Date :
Page:

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

Mathematical Model

$$y_{ij} = \mu + T_i + e_{ij} \quad [i = 1, 2, 3 \ldots, t \; ; \; j = 1, 2, 3, \ldots, r]$$

Where,

y_{ij} = Observation due to ith treatment and jth replication

μ = General Mean

T_i = Effect due to ith treatment

e_{ij} = Error due to ith treatment and jth replication

And,

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

$T_i = \bar{y}_{i.} - \bar{y}_{..}$

$e_{ij} = y_{ij} - \bar{y}_{i.}$

Then,

Total sum of square (TSS) $= \sum_i \sum_j (\bar{y}_{ij} - \bar{y}_{..})^2$

$$= \sum_i \sum_j \left(y_{ij}^2 - \frac{G^2}{N} \right)$$

Sum of square due to Treatment (SST) $= \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$

$$= \sum_i \frac{T_{i.}^2}{r} - \frac{G^2}{N}$$

Sum of square due to Error (SSE) = TSS − SST

Date : ....
Page: ....

ANOVA Table

SV	df	SS	MS	Fcal	Ftab
Treatment	$t-1$	SST	$MST = \frac{SST}{t-1}$	$F_T = \frac{MST}{MSE}$	$F_\alpha[(t-1), t(r-1)]$
Error	$t(r-1)$	SSE	$MSE = \frac{SSE}{t(r-1)}$		
Total	$rt-1$	TSS			

Q5) For the ANOVA summary table below fill in all the missing results. Also indicate your statistical decision for 4 different treatments.

SV	Df	SS	MS	F
Treatment	?	?	70	?
Error	12	590	?	
Total	?	?		

→ Soln, Given,

No. of treatments $(t) = 4$

Mean Sum of square due Treatment $(MST) = 70$

Sum of Square due to Error $(SSE) = 590$

Degree of freedom for Error $(df) = 12$

Then,

Sum of Square due to Treatment $(SST) = MST \times (t-1)$

$$SST = 70 \times 3$$
$$= 210.$$

And

Mean Sum of square due to Error $(MSE) = \frac{SSE}{t(r-1)}$

Date:
Page:

$MSE = \frac{590}{12}$

$\therefore MSE = 49.1667$

So, Total sum of square (TSS) = SST + SSE

= 210 + 590

= 800

Now,

$F_{cal} = \frac{MST}{MSE} = \frac{70}{49.1667} = 1.4237$

i.e.

SV	Df	SS	MS	F
Treatment	t-1=3	210	70	1.4237
Error	12	590	49.1667	
Total	15	800		

Q6) Complete the partially Completed ANOVA Table below.
Complete the ANOVA Table and answer the following

Source of variance	Sum of Square	Degree of freedom	Mean Sum of Square	F-value
Colums	72	?	?	2
Rows	?	?	36	?
Treatments	180	3	?	?
Error	?	6	12	
Total	?	2		

i. What design was employed?
ii. How many treatments were compared?

Date :

Page:

→ Soln

Since, The variation between Colum's, rows and treatments are studied, it is Latin Square Design (LSD).

Then,

No. of rows = No. of Columns = m

Where,

df of treatment = df of row = df of Column = m − 1.

i.e. m − 1 = 3

m = 4

So,

df of treatment = df of ~~treatment~~ row = df of Column = 3

And

Sum of square due to row (SSR) = Mean Sum of Square due to row (MSR) * df of row.

i.e. SSR = 36*3 = 108.

Similarly,

Mean sum of square due to treatment (MST)

$= \dfrac{\text{Sum of Square due to Treatment (SST)}}{\text{df of treatment}}$

$= \dfrac{180}{3} = 60$.

And,

Sum of square due to error (SSE) = Mean sum of square due to error (MSE) * df error

= 12*6

= 72

Date : ....
Page: ....

Then, Mean sum of square due to column (MSC)

$$= \frac{\text{Sum of Square due to column (SSC)}}{\text{df of Column}}$$

$$= \frac{72}{3} = 24.$$

Total sum of square (TSS) = SSC + SSR + SST + SSE
= 72 + 108 + 180 + 72
= 432

Now,

Calculation of F values.

$$F_R = \frac{MSR}{MSE} = \frac{36}{12} = 3.$$

$$F_T = \frac{MST}{MSE} = \frac{60}{12} = 5$$

Here,

Compled ANOVA Table is

Source of Variation	Sum of square	Degree of free dom	Mean sum of square	F-value
Columns	72	3	24	2
Rows	108	3	36	3
Treatments	180	3	60	5
Errors	72	6	12	
Total	432	15		

CamScanner

Date :
Page:

Q7) What do you mean by Randomized Block Design? Give mathematical model of design with meaning. Following information represent yield on using different types of treatments.

A 12	C 16	A 11	D 17
B 18	D 14	C 15	A 11
C 15	A 13	B 12	B 15
D 11	B 12	D 14	C 16

Carry out analysis of the design

→ Randomized Block Design (RBD)

- It is an experimental design used in heterogeneous material that uses all 3 principles of design namely replication, randomization and local control. Treatments are replicated along row or column and effect due to treatment and block ~~along~~ can be determined. Thus, it is a case of 2 way ANOVA.

Mathematical Model

$y_{ij} = \mu + T_i + B_j + e_{ij}$

Where,

$i = 1, 2, 3, \ldots, t$; $j = 1, 2, 3, \ldots, r$

y_{ij} = Observation due to ith treatment & jth block

μ = General mean

T_i = Effect due to ith treatment

B_j = Effect due to jth block

e_{ij} = Error due to ith treatment and jth block.

→ Sol

Sol^n

The given design is RBD.

* Problem to test

$H_0T: \mu_{1.} = \mu_{2.} = \mu_{3.} = \cdots = \mu_{i.}$

H_1T: At least one $\mu_{i.}$ is different

$H_0B: \mu_{.1} = \mu_{.2} = \mu_{.3} = \cdots = \mu_{.j}$

H_1B: At least one $\mu_{.j}$ is different.

Now,

		I	II	III	IV	$T_{i.}$
Treatments	A	12	13	11	11	47
	B	18	12	12	15	57
	C	15	16	15	16	62
	D	11	14	14	17	56
	$T_{.j}$	56	55	52	59	$G = 222$

$$\sum_i \sum_j y_{ij}^2 = 12^2+13^2+11^2+11^2+18^2+12^2+12^2+15^2+15^2+16^2+$$
$$15^2+16^2+11^2+14^2+14^2+17^2$$
$$= 3156$$

Here,

Treatments $(t) = 4$ (A, B, C, D)

Replication block $(r) = 4$

$G = \sum_i T_{i.} = \sum_j T_{.j} = 222$

$N = rt = 4 \times 4 = 16$

Then,

Date : ....

Page: ....

Total sum of square (TSS) $= \sum_{ij} y_{ij}^2 - \frac{G^2}{N}$

$= 3156 - \frac{222^2}{16} = 75.75$

Sum of square due to Treatment (SST) $= \sum_i \frac{T_{i.}^2}{r} - \frac{G^2}{N}$

$= \frac{47^2+57^2+62^2+56^2}{4} - \frac{222^2}{16} = 29.25$

Sum of square due to block (SSB) $= \sum_j \frac{T_{.j}^2}{t} - \frac{G^2}{N}$

$= \frac{56^2+55^2+52^2+59^2}{4} - \frac{222^2}{16}$

$= 6.25$

Sum of square due to Error (SSE) $= TSS - SST - SSB$

$= 75.75 - 29.25 - 6.25$

$= 40.25$

Degree of freedom (df)

df treatment $= t-1 = 4-1 = 3$

df block $= r-1 = 3$

df error $= (r-1)(t-1) = 3 \times 3 = 9$

df Total $= N-1 = 16-1 = 15$

ANOVA Table

let, $\alpha = 5\% = 0.05$ be level of Significance.

Date:
Page:

SV	df	SS	MS	Fcal
Treatment	3	29.25	$MST = SST/t-1 = 9.75$	$F_T = MST/MSE = 2.1801$
Block	3	6.25	$MSB = SSB/t-1 = 1.0833$	$F_B = MSB/MSE = 0.4658$
Error	9	40.25	$MSE = SSE/(t-1)(r-1) = 4.4722$	
Total	15	75.75		

Ftab

$F_{0.05}(3,9) = 3.862$.

Decision

$F_T = 2.1801 < F_\alpha\{(t-1),(t-1)(r-1)\} = F_{0.05}(3,9) = 3.862$

Accept H_{0T} at $\alpha = 0.05$

$F_B = 0.4658 < F_{0.05}(3,9) = 3.862$

Accept H_{0B} at $\alpha = 0.05$.

Conclusion

Treatments are insignificant

Blocks are insignificant.

Q8) Everyday is generally considered as either sunny or rainy. A sunny day is followed by another sunny day with probability 0.6 whereas a rainy day is followed by a sunny day with probability 0.3. Suppose it rains on Monday. Make forecasts for Tuesday & wednesday.

→ Soln,

Date : ....

Page: ....

let, Sunny day is denoted by 1
Rainy day is denoted by 2

$$
\begin{array}{c|cc}
\{\text{From (आउने)}\} & 1 & 2 \ \{\text{To (जाने)}\} \\
1 & 0.6 & 0.4 \\
2 & 0.5 & 0.5
\end{array}
$$

Since, It rained on Monday,

- For Tuesday
  - Probability of sunny day $(P_{21}) = 0.5 = 5\%$
  - Probability raing day $(P_{22}) = 0.5 = 50\%$

- For wednesday
  The required probability matrix is

$$P^{(2)} = P * P$$

$$= \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.56 & 0.44 \\ 0.55 & 0.45 \end{bmatrix}$$

i.e.

Probability of Sunny day $P^{(2)}_{21} = 0.55 = 55\%$

Probability of rainy day $P^{(2)}_{22} = 0.45 = 45\%$

Hence,

There is 50% probability of sunny day on Tuesday
" " 50% " " rainy " "
" 55% " " Sunny " wednesday
" 45% " " Rainy " "

Date :

Page:

10) In Some town, each day is either sunny or rainy. A sunny day is followed by another sunny day with probability 0.7, whereas a rainy day is followed by a sunny day with probability 0.4. Weather Conditions in this problem represent a homogeneous Markov Chain with 2 states: state 1 = "sunny" and state 2 = "rainy". Transition probability matrix of sunny and rainy days is given below.

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

Compute the probability of sunny days and rainy days using the steady-state equations for this Markov Chain.

→ Soln,

We are given

State 1 = "sunny" and state 2 = "rainy".

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \end{array}$$

let,

π_1 and π_2 represent steady state probabilities for sunny days and rainy days.

Then,

$\pi = [\pi_1 \quad \pi_2]$ is their matrix representation

Now,

We know,

$$\pi P = \pi$$

$$\text{or, } [\pi_1 \quad \pi_2] \cdot \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = [\pi_1 \quad \pi_2].$$

Date :

Page:

or, $\begin{bmatrix} 0.7\pi_1 + 0.4\pi_2 & 0.3\pi_1 + 0.6\pi_2 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$

Equating corresponding elements of matrix.

$0.7\pi_1 + 0.4\pi_2 = \pi_1$ —— (1)

$0.3\pi_1 + 0.6\pi_2 = \pi_2$ —— (2)

Also,

We know,

$$\sum_i \pi_i = 1$$

i.e. $\pi_1 + \pi_2 = 1$ ⟶ (3)

Here,

Eqn (1) can be written as,

$0.7\pi_1 + 0.4\pi_2 = \pi_1$

$0.7\pi_1 + 0.4(1-\pi_1) = \pi_1$ [Using eqn (3)]

$0.3\pi_1 - \pi_1 = -0.4$

$\pi_1 = \frac{0.4}{0.7} = 0.5714$

Then,

From eqn (3)

$\pi_2 = 1 - \pi_1 = 1 - 0.5714 = 0.4286$

Hence,

The steady state probability for sunny day is 0.5714 and for rainy day is 0.4286.