ANOVA —ONE WAY CLASSIFICATION

PROBLEM 1. Three processes A, B and C are tested to see whether their outputs are equivalent. The following observations of output are made :

A	10	12	13	11	10	14	15	13
B	9	11	10	12	13			
C	11	10	15	14	12	13		

Carry out the analysis of variance and state your conclusions.

Simplified Formula for calculations

In the practical situations, we can use the following formula for easy of calculations.

Let

$$G = \text{Grand total} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$$

$$RSS = \text{Raw Sum of Squares} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2$$

$$N = \sum_{i=1}^{k} n_i$$

$$TSS = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \bar{y}_{..}^2 - 2 \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij} \bar{y}_{..}$$

$$= RSS + N \bar{y}_{..}^2 - 2 \bar{y}_{..} N \bar{y}_{..} = RSS - N \bar{y}_{..}^2$$

$$= RSS - N \left(\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}}{N} \right)^2 = RSS - N \left(\frac{G}{N} \right)^2$$

$$TSS = RSS - \frac{G^2}{N}$$

$\frac{G^2}{N}$ is called Correction Factor (C.F.)

$$\therefore \quad TSS = RSS - CF$$

$$SST = \sum_{i=1}^{k} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{k} n_i \bar{y}_{i.}^2 + \sum_{i=1}^{k} n_i \bar{y}_{..}^2 - 2\bar{y}_{..} \sum_{i=1}^{k} n_i \bar{y}_{i.}$$

$$= \sum_{i=1}^{k} n_i \bar{y}_{i.}^{2} + N\bar{y}_{..}^{2} - 2\bar{y}_{..} N\bar{y}_{..}$$

$$= \sum_{i=1}^{k} n_i \left(\frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \right)^2 - N\bar{y}_{..}^{2} = \sum_{i=1}^{k} n_i \left(\frac{T_{i.}}{n_i} \right)^2 - \frac{G^2}{N}$$

$$SST = \sum_{i=1}^{k} \frac{T_{i.}^{2}}{n_i} - C.F. \quad \text{and} \quad SSE = TSS - SST$$

...analysis of variance and state your conclusions.

SOLUTION

Null Hypothesis

H_0: Outputs of three processes A, B and C are equal.

i.e., $H_0: \mu_1 = \mu_2 = \mu_3$

Processes	*Observations*								$T_{i.}$	$\frac{T_{i.}^2}{n_i}$	$\sum_j y_{ij}^2$
A	10	12	13	11	10	14	15	13	98	1200.5	1224
B	9	11	10	12	13				55	605	615
C	11	10	15	14	12	13			75	937.5	955
								Totals	$\sum_{i=1}^{k} T_{i.} = 228$	$\sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} = 2743$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = 2794$

$$G = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} T_{i.} = 228$$

$$RSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = 2794$$

Correction factor

$$C.F. = \frac{G^2}{N} = \frac{(228)^2}{19} = 2736$$

$$TSS = RSS - CF$$

$$= 2794 - 2736 = 58$$

$$SST = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - C.F.$$

$$= 2743 - 2736 = 7$$

$$SSE = TSS - SST$$

$$= 58 - 7 = 51$$

ANOVA TABLE

Source of Variation	*S.S*	*d.f*	*M.SS*	*F-ratio (Variation ratio)*
Between Process (Treatments)	7	$(k-1=3-1)$ 2	$\frac{7}{2}=3.5$	$F=\frac{3.5}{3.1875}=1.098$
Within process (error)	51	16 $(N-k=19-3)$	$\frac{51}{16}=3.1875$	$\sim F_{(2,16)}$
Total	58	18 $(N-1=19-1)$		

Tabulated value of F at 5% level is

$$F_{(5\%,(k-1,N-k))}=F_{(5\%,2,16)}=3.63$$

$$F_{cal}=1.098<F_{tab}=3.63$$

Conclusion

Hence H_0 may be accepted at 5% level of significance *i.e.,* outputs of three processes A, B and C are equal.

ANOVA–TWO WAY CLASSIFICATION

PROBLEM 1. There are four doctors, if they wish to test the five medicines, they applied these five treatments *i.e.*, medicines on four patients each and the reading were given below : (1% L.o.S)

	Observations of Treatments (Medicines)				
	A	**B**	**C**	**D**	**E**
1	12	16	18	21	24
2	16	25	20	23	28
3	14	20	23	16	20
4	15	24	23	25	36

Simplified Formulae for Calculations

$$G = \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij} \text{ is grand total}$$

$$C.F. = \frac{G^2}{N} \text{ is correction factor}$$

$$N = hk$$

$$RSS = \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij}^2 \text{ is Raw sum of squares}$$

$$TSS = RSS - CF$$

$$SST = \frac{1}{h} \sum_{i=1}^{k} T_{i.}^2 - CF$$

$$SSV = \frac{1}{k} \sum_{j=1}^{h} T_{.j}^2 - CF$$

$$SSE = TSS - SST - SSV$$

SOLUTION. Null Hypothesis

H_T (or) H_M : There is no significance difference between the medicines

i.e., $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu_{5.}$

H_V (or) H_D : There is no significance difference between the doctors.

i.e., $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4}$

Doctors	*Medicines (Treatments)*					$T_{.j}$	$T_{.j}^2$
	A	*B*	*C*	*D*	*E*		
1	12	16	18	21	24	91	8281
2	16	25	20	23	28	112	12544
3	14	20	23	16	20	93	8649
4	15	24	23	25	36	123	15129
$T_{i.}$	57	85	84	85	108	G = 419	44603
$T_{i.}^2$	3249	7225	7056	7225	11664	36419	

$$G = 419, \quad N = 20, \quad k = 5, h = 4$$

$$\text{C.F.} = \frac{G^2}{N} = \frac{(419)^2}{20} = 8778.05$$

$$\text{RSS} = \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij}^2 = 9367$$

$$\text{SST} = \text{SSM} = \frac{1}{h} \sum_{i=1}^{k} T_{i.}^2 - \text{CF}$$

$$= \frac{36419}{4} - 8778.05 \simeq 326.7$$

$$\text{SSV} = \text{SSD} = \frac{1}{k} \sum_{j=1}^{h} T_{.j}^2 - \text{CF}$$

$$= \frac{44603}{5} - 8778.05 = 142.55$$

$$\text{TSS} = \text{RSS} - \text{CF}$$

$$= 9367 - 8778.05 = 588.95$$

$$\text{SSE} = \text{TSS} - \text{SST} - \text{SSV}$$

$$= 588.95 - 326.7 - 142.55 = 119.7$$

ANOVA Table

Source of Variation	*Sum of Squares SS*	*d.f.*	*M.SS*	*F-ratio*
Treatment (medicines)	326.7	4 $(k-1)$	$\frac{326.7}{4} = 81.675$	F_T (or) $F_M = \frac{81.675}{9.9745} = 8.19 \sim F_{(4, 12)}$
Varieties (Doctors)	142.55	3 $(h-1)$	$\frac{142.55}{3} = 47.52$	F_V (or) $F_D = \frac{47.52}{9.975} = 4.76 \sim F_{(3, 12)}$
Error	119.7	12	$\frac{119.7}{12} = 9.975$	
Total	588.95	19 $(N-1)$		

Conclusion

(1) For treatments (medicines)

$F_T = 8.19, \quad F_{(1\%, (4, 12))} = 5.41$

$F_{T\,(calculated)} > F_{tabulated}$

$\therefore$ H_T may be rejected *i.e.,* the effects of medicines differ significantly.

(2) For varieties (Doctors)

$F_V = 4.76,$

$F_{(1\%, (3, 12))} = 5.95$

$F_{V\,(Calculated)} < F_{(1\%, (3,12))}$ tabulated

Hence we may accept H_V.

i.e., there is no significance difference between doctors at 1% level of significance.

Completely Randomised Design (CRD)

ANOVA—CRD

PROBLEM 1. The following figures related to the production of wheat in kgs of yield by applying 3 treatments A, B, C in 12 plots.

A	14	16	18	—	—
B	14	13	15	22	—
C	18	16	19	19	20

Is their any significant difference in the production of 3 varieties ?

1. Aim : To analyse the given data by using CRD.

Formula and Procedure :

Hull Hypothesis H_0 : The means of various treatments effects are homogeneous.

i.e., $\mu_1 = \mu_2 = \mu_3 = \mu$ (or) $\alpha_1 = \alpha_2 = \alpha_3 = 0$

Alternative Hypothesis H_1 : The means of various treatment effects are not homogeneous

i.e., $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu$ (or) $\alpha_1 \neq \alpha_2 \neq \alpha_3 \neq 0$

Raw sum of squares = RSS = $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2$

Grand Total = G = $\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} = \sum_{i=1}^{k} T_i$

N = Total number of observations.

Correction factor (CF) = $\frac{G^2}{N}$

Total sum of squares TSS = RSS – CF

Sum of squares due to treatments

$$SST = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - CF$$

Sum of squares due to error

$$SSE = TSS - SST$$

Calculations :

Null hypothesis H_0 **:** The mean effect of production of wheat is homogeneous w.r.t. to the 3 treatments A, B, C

i.e. $t_A = t_B = t_C$

Alternative Hypothesis H_1 **:** The mean effect of production of wheat is not homogeneous w.r.t. to the 3 treatments A, B, C

i.e. $t_A \neq t_B \neq t_C$

						n_i	T_i	T_i^2/n_i	$\sum x_{ij}^2$
A	14	16	18	–	–	3	48	768	776
B	14	13	15	22	–	4	64	1024	1074
C	18	16	19	19	20	5	92	1692.8	1702
						N = 12	$\sum T_i = 204$	$\sum T_i^2/n_i = 3484.8$	3552

Here k = Number of treatments

$k = 3$, N = 12

Grand total $\sum_i T_i = 204$

$$CF = \frac{G^2}{N} = \frac{(204)^2}{12} = 3408$$

Raw sum of squares = $\sum x_{ij}^2 = 3552$

Total sum of squares TSS = RSS – CF = 3552 – 3468 = 84

Sum of squares due to treatments, $SST = \sum_i \left(\frac{T_i^2}{n_i}\right) - CF$

$= 3484.8 - 3468 = 16.8$

Sum of squares due to error, SSE $= TSS - SST$

$= 84 - 16.8 = 67.2$

ANOVA Table for CRD

S.V	*D.F*	*SS*	*MSS*	*VR*
Treatments	$k - 1 = 3 - 1 = 2$	SST = 16.8	$s_t^2 = 8.4$	$F_t = \frac{s_t^2}{s_E^2} = 1.1249$
Error	$N - k = 12 - 3 = 9$	SSE = 67.2	$s_E^2 = 7.4667$	–
Total	$N - 1 = 12 - 1 = 11$	TSS = 84	–	–

F_t table value at (2, 9) *d.f.* 5% level of significances = 4.26

Since $F_t <$ F table value. So we accept H_0. $1.1249 < 4.26$ ✓

Inference :

The mean effect of production of wheat is homogeneous with respect to 3 treatments A, B, C *i.e.* $t_A = t_B = t_C$

Randomised Block Design (RBD)

PROBLEM 1. 3 varieties of coal were analysed by 4 chemists and the ash content in the varieties was found to be as follows :

Varieties	*Chemists*			
	1	*2*	*3*	*4*
A	8	5	5	7
B	7	6	4	4
C	3	6	5	4

Carry out analysis of R.B.D.

PROBLEM 2. Analyse the following randomised block design after estimating the missing value.

Treatments	*Blocks*			
	1	*2*	*3*	*4*
1	18	27	22	20
2	20	15	21	11
3	15	21	14	19
4	32	20	—	22
5	20	25	26	24

1. Aim : To analyse the data by using RBD.

Formula and Procedure :

Null Hypothesis ***(1)*** H_{01}: $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu$ (or) $\alpha_1 = \alpha_2 = \alpha_3 = 0$

(2) H_{02}: $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu$ (or) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

i.e., the treatment and blocks effects are homogeneous.

Grand Total, $G = \sum_{i=1}^{k} T_{i.} = \sum_{j=1}^{h} T_{.j}$

Correction factor, $CF = \frac{G^2}{N}$, $N = kh$

Total sum of squares, TSS = RSS – CF

Raw Sum of squares, $RSS = \sum_{i=1}^{k} \sum_{j=1}^{h} x_{ij}^2$

Sum of squares due to treatments

$$SST = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - CF$$

PROBLEM 1. 3 varieties of coal were analysed by 4 chemists and the ash content in the varieties was found to be as follows :

Varieties	*Chemists*			
	1	*2*	*3*	*4*
A	8	5	5	7
B	7	6	4	4
C	3	6	5	4

Carry out analysis of R.B.D.

1. Aim : To analyse the data by using RBD.

Formula and Procedure :

Null Hypothesis ***(1)*** H_{01} : $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu$ (or) $\alpha_1 = \alpha_2 = \alpha_3 = 0$

(2) H_{02} : $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu$ (or) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

i.e., the treatment and blocks effects are homogeneous.

Grand Total, $G = \sum_{i=1}^{k} T_{i.} = \sum_{j=1}^{h} T_{.j}$

Correction factor, $CF = \frac{G^2}{N}$, $N = kh$

Total sum of squares, TSS = RSS – CF

Raw Sum of squares, $RSS = \sum_{i=1}^{k} \sum_{j=1}^{h} x_{ij}^2$

Sum of squares due to treatments

$$SST = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - CF$$

Sum of squares due to varieties

$$SSV = \sum_{j=1}^{h} \frac{T_j^2}{n_j} - CF$$

Sum of squares due to error

$$SSE = TSS - SST - SVV$$

Calculations :

Null Hypothesis is H_{01} **:** The mean effect of various chemists are homogeneous

H_{02} **:** The mean effect of various varieties are homogeneous

i.e. $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu$

and $\mu_{.A} = \mu_{.B} = \mu_{.C} = \mu$

Alternative Hypothesis

H_{11} : The mean effect of various chemists are not homogeneous

H_{22} : The mean effect of various varieties are not homogeneous

i.e. $\mu_{1.} \neq \mu_{2.} \neq \mu_{3.} \neq \mu_{4.} \neq \mu$

and $\mu_{.A} \neq \mu_{.B} \neq \mu_{.C} \neq \mu$

Varieties	*Chemists*							
	1	*2*	*3*	*4*	n_i	T_i	T_i^2/n_i	$\sum x_{ij}^2$
A	8	5	5	7	4	25	156.25	163
B	7	6	4	4	4	21	110.25	117
C	3	6	5	4	4	18	81	86
n_j	3	3	3	3	N = 12	G = 64	347.5	$\sum\sum x_{ij}^2$
T_j	18	17	14	15	G = 64			
T_j^2/n_j	108	96.333	65.333	75	344.66			

Here k = Number of treatments = 4

h = Number of varieties = 3

$$N = kh = 12$$

$$CF = \frac{G^2}{N} = \frac{(64)^2}{12} = 341.333$$

$$RSS = \sum\sum x_{ij}^2 = 366$$

Total sum of squares $TSS = RSS - CF = 366 - 341.333 =- 24.667$

Sum of squares due to treatments, $SST = \sum\left(\frac{T_i^2}{n_i}\right) - CF$

$$= 347.5 - 341.33 = 6.1667$$

Sum of squares due to the varieties, $SSV = \sum\left(\frac{T_j^2}{n_j}\right) - CF$

$$= 344.66 - 341.33 = 3.333$$

Sum of squares due to error, $SSE = TSS - SST - SSV$

$$= 24.667 - 6.1667 - 3.333$$

$$SSE = 15.1667$$

ANOVA Table for RBD

SV	*D/F*	*SS*	*MSS*	*VR*
Treatments	$k-1 = 4-1 = 3$	$SST = 6.1667$	$s_t^2 = 2.0556$	$F_t = \frac{s_t^2}{s_E^2} = 0.8132$
Varieties	$h-1 = 3-1 = 2$	$SSV = 3.3333$	$s_V^2 = 1.6666$	$F_V = \frac{s_V^2}{s_E^2} = 0.6593$
Error	$(k-1)(h-1) = 6$	$SSE = 15.1667$	$s_E^2 = 2.52778$	
Total	$kh - 1 = 11$	$TSS = 24.667$	—	—

F_t calculated value for chemists = 0.8132

F_V calculated value for varieties = 0.6593

Table value of F (3, 6) *d.f.* at 5% level of significance = 4.76

Table value of F (2, 6) *d.f.* at 5% level of significance = 5.14

Since F_t calculated value < F table value. So we accept H_0.

Since F_V calculated value < F table value. So we accept H_0.

Inference :-

The mean effect of various chemists are homogeneous

The mean effect of various varieties are homogeneous

i.e. $\mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.} = \mu$

and $\mu_{.A} = \mu_{.B} = \mu_{.C} = \mu$

Latin Square Design (LSD)

PROBLEM 1. Study the effectiveness of 3 teaching methods A, B, C from the achievement scores given below tabulated age and aptitude wise.

Aptitude	*Age*					
	Young		*Middle*		*Old*	
Low	A	82	B	87	C	80
Middle	B	92	C	82	A	81
High	C	90	A	83	B	88

Aim : To analyse the data by using LSD.

Formula and Procedure :

Null Hypothesis

1. $H_{0\alpha} : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$
2. $H_{0\beta} : \beta_1 = \beta_2 = \ldots = \beta_m = 0$
3. $H_{0\gamma} : \gamma_1 = \gamma_2 = \ldots = \gamma_m = 0$

i.e., the row effects, column effects and treatment effects are homogeneous.

Grand Total = $G = \sum\sum x_{ijk}$

$N = m^2$

Correction factor $C.F = \frac{G^2}{N} = \frac{G^2}{m^2}$

Raw sum of squares = RSS = $\sum\sum x_{ijk}^2$

Total sum of squares = TSS = RSS – CF

Sum of square due to rows

$$SSR = \frac{1}{m}\sum_{i=1}^{m} R_i^2 - CF$$

Sum of square due to columns

$$SSC = \frac{1}{m}\sum_{j=1}^{m} C_j^2 - CF$$

Sum of squares due to treatments

$$SST = \frac{1}{m}\sum_{k=1}^{m} T_k^2 - CF$$

Sum of squares due to error

$$SSE = TSS - SSR - SSC - SST$$

1. Calculations :

Null Hypothesis H_0 : The mean effect of different aptitudes are homogeneous. The mean effect of different age groups are homogeneous and the mean effect of various teaching methods are homogeneous

i.e. $H_{0R} : R_1 = R_2 = R_3$

$H_{0C} : C_1 = C_2 = C_3$

$H_{0t} : t_A = t_B = t_C$

Alternate Hypothesis H_1 : The mean effect of different aptitudes are not homogeneous. The mean effect of different age groups are not homogeneous and the mean effect of various teaching methods are not homogeneous.

i.e., $H_{1R} : R_L \neq R_2 \neq R_3$

$H_{1C} : C_1 \neq C_2 \neq C_3$

$H_{1t} : t_A \neq t_B \neq t_C$

	A	***B***	***C***	***Total***
T_k	246	267	252	765
T_k^2	60516	71289	63504	195309

Aptitude	***Age***					
	Young	***Middle***	***Old***	R_i	R_i^2	$\sum x_{ijk}^2$
Long	82	87	80	249	62001	20693
Middle	92	82	81	255	65025	21749
High	90	83	88	261	68121	22733
C_j	264	252	249	765	195147	65175
C_j^2	69696	63504	62001	95201		

$$G = \text{Grand total} = 765$$

$$N = 3 \times 3 = 9$$

$$CF = \frac{G^2}{N} = \frac{(765)^2}{9} = 65025$$

$$RSS = \text{Raw sum of squares} = \sum\sum x_{ijk}^2 = 65175$$

$$TSS = RSS - CF$$

$$= 65175 - 65025 = 150$$

$$SSR = \frac{1}{m}\sum_{i=1}^{m} R_i^2 = CF = \frac{195147}{3} - 65025 = 24$$

$$SSC = \frac{1}{m}\sum_{j=1}^{m} C_j^2 - CF = \frac{195201}{3} - 65025 = 42$$

$$SST = \frac{1}{m}\sum_{k=1}^{m} T_k^2 - CF = \frac{195309}{3} - 65025 = 78$$

$$SSE = TSS - SST - SSR - SSC$$

$$= 150 - 78 - 24 - 42 = 6$$

ANOVA Table for LSD

S.V	*DF*	*SS*	*MSS*	*VR*	*F-table value*
Rows	$3 - 1 = 2$	SSR = 24	$s_R^2 = 12$	$F_R = 4$	F(2, 2) d.f. at 5% = 19.0
Columns	$3 - 1 = 2$	SSC = 42	$s_C^2 = 21$	$F_C = 7$	F(2, 2) d.f. = 19.00
Treatments	$3 - 1 = 2$	SST = 78	$s_t^2 = 39$	$F_t = 13$	F(2, 2) d.f. = 19.00
Error	$(3 - 1)(3 - 2) = 2$	SSC = 6	$s_E^2 = 3$	—	—
Total	$3^2 - 1 = 8$	TSS = 150	—	—	—

Inference :

Ft-calculated value < F table value at 5% level of significance. So we accept H_0. *i.e.,* the mean effect of various aptitudes are homogeneous. The mean effect of different age groups are homogeneous and the mean effect of various teaching methods are homogeneous.

ANOVA—RBD with and without one Missing Observation

PROBLEM 1. 3 varieties of coal were analysed by 4 chemists and the ash content in the varieties was found to be as follows :

Varieties	*Chemists*			
	1	*2*	*3*	*4*
A	8	5	5	7
B	7	6	4	4
C	3	6	5	4

Carry out analysis of R.B.D.

PROBLEM 2. Analyse the following randomised block design after estimating the missing value.

Treatments	*Blocks*			
	1	*2*	*3*	*4*
1	18	27	22	20
2	20	15	21	11
3	15	21	14	19
4	32	20	—	22
5	20	25	26	24

ANOVA—RBD with and without one Missing Observation

PROBLEM 2. Analyse the following randomised block design after estimating the missing value.

Treatments	*Blocks*			
	1	***2***	***3***	***4***
1	18	27	22	20
2	20	15	21	11
3	15	21	14	19
4	32	20	—	22
5	20	25	26	24

2. Aim : To carryout statistical analysis of RBD with missing observation.

Formula and Procedure :

The missing observation is calculated by

$$x = \frac{k\,T^{1}_{i.} + h\,T^{1}_{.j} - G^{1}}{(h-1)(k-1)}$$

where $T^1_{i.}$ is sum of known observations due to the *i*th treatment

$T^1_{.j}$ is sum of known observations due to *j*th block.

G^1 is sum of all known observations

h = number of blocks

k = number of treatments.

The statistical analysis of RBD is as usual explained in problem 1.

Calculations :

Treatments	*Blocks* 1	2	3	4	$T_{i.}$
1	18	27	22	20	87
2	20	15	21	11	67
3	15	21	14	19	69
4	32	20	x (say)	22	$74 + x = 99.83$
5	20	25	26	24	95
$T_{.j}$	105	108	$83 + x = 108.83$	96	$392 + x = 417.83$

$$T^1_{i.} = 74, \quad T^1_{.j} = 83, \quad G^1 = 392$$

$$h = 4, \quad k = 5$$

$$\therefore \quad x = \frac{k\,T^1_{i.} + h\,T^1_{.j} - G^1}{(h-1)(k-1)}$$

$$= \frac{5 \times 74 + 4 \times 83 - 392}{(4-1)(5-1)} = 25.83$$

$$\therefore \quad T_{i.} = T^1_{i.} + x = 74 + 25.83 = 99.83$$

$$T_{.j} = T^1_{.j} + x = 83 + 25.83 = 108.83$$

$$G = G^1 + x = 392 + 25.83 = 417.83$$

$$CF = \frac{G^2}{N} = \frac{(417.83)^2}{20} = 8729.10$$

$$RSS = \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij}^2$$

$$= (18)^2 + (27)^2 + \ldots + (24)^2 = 9199.19$$

$$TSS = RSS - CF$$

$$= 9199.19 - 8729.1 = 370.09$$

$$SST = \frac{1}{h} \sum_{i=1}^{k} T_{i.}^2 - CF$$

$$= \frac{(87)^2 + (67)^2 + (69)^2 + (99.83)^2 + (95)^2}{4} - 8729.1$$

$$= 223.41$$

$$SSB = \frac{1}{k} \sum_{j=1}^{h} T_{.j}^2 - CF$$

$$= \frac{105^2 + 108^2 + 108.83^2 + 96^2}{5} - 8729.1$$

$$= 20.69$$

$$\therefore \quad SSE = TSS - SST - SSB$$

$$= 370.09 - 223.41 - 20.69 = 125.99$$

Null Hypothesis

$$H_T : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$

$$H_B = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

ANOVA Table for RBD

Source of Variation	*SS*	*d.f.*	*MSS*	*F-ratio*
Treatments	223.41	4	$\frac{223.41}{4} = 55.85$	$F_T = \frac{55.85}{10.50} = 5.32 \sim F_{(4, 12)}$
Blocks	20.69	3	$\frac{20.69}{3} = 6.90$	$F_B = \frac{10.50}{6.9} = 1.52 \sim F_{(12, 3)}$
Error	125.99	12	$\frac{125.99}{12} = 10.50$	
Total	370.09	19		

Comparison and Conclusion

(1) ***For Treatments***

$$F_T = 5.32$$

$$F_{(5\%, (4, 12))} = 3.26$$

$$\Rightarrow \quad F_T > F_{(5\%, (4, 12))}$$

$\therefore$ H_T may be rejected.

Treatments effects are not homogeneous.

(2) ***For Blocks***

$$F_B = 1.52,$$

$$F_{(5\%, (12, 3))} = 8.74$$

$\therefore$ $F_B < F_{(5\%, (12, 3))} \Rightarrow$ H_B may be accepted.

i.e., Blocks effects are homogeneous.

Inference :

1. The missing observation is $x = 25.83$
2. The treatment effects are not homogeneous
3. The block effects are homogeneous.

Analysis of LSD With and Without one Missing Observation

PROBLEM 3. Analyse the following Latin square design after estimating the missing value.

A 12	C 19	B 10	D 8
C 18	B 12	D 6	A —
B 22	D 10	A 5	C 21
D 12	A 7	C. 27	N 17

Aim : To analyse the data by using LSD.

Formula and Procedure :

Null Hypothesis

1. $H_{0\alpha} : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$
2. $H_{0\beta} : \beta_1 = \beta_2 = \ldots = \beta_m = 0$
3. $H_{0\gamma} : \gamma_1 = \gamma_2 = \ldots = \gamma_m = 0$

i.e., the row effects, column effects and treatment effects are homogeneous.

Grand Total = $G = \sum\sum x_{ijk}$

$N = m^2$

Correction factor $C.F = \frac{G^2}{N} = \frac{G^2}{m^2}$

Raw sum of squares = RSS = $\sum\sum x_{ijk}^2$

Total sum of squares = TSS = RSS − CF

Sum of square due to rows

$$SSR = \frac{1}{m}\sum_{i=1}^{m} R_i^2 - CF$$

Sum of square due to columns

$$SSC = \frac{1}{m}\sum_{j=1}^{m} C_j^2 - CF$$

Sum of squares due to treatments

$$SST = \frac{1}{m}\sum_{k=1}^{m} T_k^2 - CF$$

Sum of squares due to error

$$SSE = TSS - SSR - SSC - SST$$

3. Calculations :

Rows	*Columns*				*Row Totals*
	I	*II*	*III*	*IV*	R_i
I	A 12	C 19	B 10	D 8	49
II	C 18	B 12	D 6	A –	$36 + x = 38$
III	B 22	D 10	A 5	C 21	58
IV	D 12	A 7	C 27	B 17	63
Column Totals C_j	64	48	48	$46 + x = 48$	$206 + x =$ 208

Treatment totals (T_k)

$$A = 24 + x,\ B = 61,\ C = 85,\ D = 36$$
$$= 24 + 2 = 26$$

$\therefore$ $R = 36,\ C = 46,\ T = 24,\ S = 206,$

$m = 4$

$\therefore$ Missing value

$$x = \frac{m\,(R + C + T) - 2S}{(m-1)\,(m-2)}$$

$$x = \frac{4\,(36 + 46 + 24) - 2 \times 206}{(4-1)\,(4-2)} = 2$$

$$G = 208$$

$$CF = \frac{G^2}{m^2} = \frac{(208)^2}{4^2} = 2704$$

$$RSS = \sum y_{ijk}^2$$
$$= (12)^2 + (19)^2 + \ldots + (17)^2 = 3438$$

$$TSS = RSS - CF$$
$$= 3438 - 2704 = 734$$

$$SSR = \frac{1}{m}\sum_{i=1}^{m} R_i^2 - CF$$
$$= \frac{(49)^2 + (38)^2 + (58)^2 + (63)^2}{4} - 2704$$
$$= 2794.5 - 2704 = 90.5$$

$$SSC = \frac{1}{m}\sum_{i=1}^{m} C_j^2 - CF$$

$$= \frac{(64)^2 + (48)^2 + (48)^2 + (48)^2}{4} - 2704 = 48$$

$$SST = \frac{1}{m}\sum_{k=1}^{m} T_k^2 - CF$$

$$= 525.5$$

$$\therefore \quad SSE = TSS - SSR - SSC - SST$$

$$= 734 - 90.5 - 48 - 525.5 = 70$$

ANOVA Table for LSD

Source of Variation	*S.S*	*d.f.*	*MSS*	*F-ratio*
Rows	90.5	3	$\frac{90.5}{3} = 30.17$	$F_R = \frac{30.17}{11.67} = 2.59 \sim F_{(3,6)}$
Columns	48	3	$\frac{48}{3} = 16$	$F_C = \frac{16}{11.67} = 1.37 \sim F_{(3,6)}$
Treatments	525.5	3	$\frac{525.5}{3} = 175.17$	$F_T = \frac{175.17}{11.67} = 15.01 \sim F_{(3,6)}$
Error	70	6	$\frac{70}{6} = 11.67$	–
Total	734	15		–

Comparison and Conclusion

1. ***Rows***

$F_R = 2.59$

$F_{(5\%, (3,6))} = 4.76$

$F_R < F_{(5\%1, (3,6))} \Rightarrow H_0$ may be accepted.

i.e., row effects are homogeneous.

2. ***Columns***

$F_C = 1.37$

$\therefore \quad F_C < F_{(5\%, (3,6))} \Rightarrow H_0$ may be accepted.

i.e., column effects are homogeneous.

3. ***Treatments***

$F_T = 15.01$

$\therefore \quad F_T > F_{(5\%, (3,6))} \Rightarrow H_0$ is rejected.

i.e., treatment effects are significant.

Inference :

1. The missing observation is $x = 2$
2. Row and column effects are homogeneous.
3. The treatment effects are not same.

2^2-FACTORIAL EXPERIMENT

Example 2 *Below is given the plan and yields of a 2^2 factorial experiment conducted in CRD. Analyse the design and give your comments:*

(1) 20	*a* 28	*a* 24	*b* 10
ab 23	*b* 11	*ab* 22	(1) 17
a 24	*b* 15	*ab* 21	(1) 19

SOLUTION Arranging the observations in one way classification, we proceed as follows:

Treatment Combination			Total	
(1)	20	17	19	56
a	28	24	24	76
b	10	11	15	36
ab	23	22	21	66
Total				234

$$\text{Correction factor} = \frac{G^2}{2^2 \times r} = \frac{234 \times 234}{4 \times 3} = 4563$$

$$\sum_i \sum_j x_{ij}^2 = (20^2 + 17^2 + 19^2) + (28^2 + 24^2 + 24^2) + (10^2 + 11^2 + 15^2) + (23^2 + 22^2 + 21^2)$$

$$= 1050 + 1936 + 446 + 1454 = 4886$$

$$TSS = 4886 - 4563 = 323$$

To obtain the sum of squares *SSA, SSB, SSAB*, we can use Yates method:

Yates Method :

Total Response	(1)	(2)	Divisor	SS
[1] 56	$56 + 76 = 132$	$132 + 102 = 234$	$4r$	$\frac{(234)^2}{12} = CF$
[*a*] 76	$36 + 66 = 102$	$20 + 30 = 50$	$4r$	$\frac{50^2}{12} = 208.33$
[*b*] 36	$76 - 56 = 20$	$102 - 132 = -30$	$4r$	$\frac{(-30)^2}{12} = 75.00$
[*ab*] 66	$66 - 36 = 30$	$30 - 20 = 10$	$4r$	$\frac{(10)^2}{12} = 8.33$
			Total	291.66

$$\therefore \quad SSE = TSS - (SSA + SSB + SSAB) = 323 - (208.33 + 75.00 + 8.33)$$

$$= 323 - 291.66 = 31.34$$

Analysis of Variance Table

Source of Variation	d.f.	*SS*	*MSS*	*F*	$F_{0.01}(1, 8)$
A	1	208.33	208.33	53.15**	11.26
B	1	75.00	75.00	19.13**	
AB	1	8.23	8.33	2.13	
Error	8	31.34	3.92	—	
Total	11	323.00	—		

**Main effects *A* and *B* both are significantly different. With *r* observations per cell we proceed as follows:

	Factor B		Total
	1	2	
Factor A 1	20, 17, 19 = 56	10, 11, 15 = 36	92
Factor A 2	28, 24, 24 = 76	23, 22, 21 = 56	142
Total	132	102	234

Let y_{ijk} represent the observaion taken under the i^{th} level of factor A and the j^{th} level of B in the k^{th} replicate, $i = 1, 2,\ j = 1, 2,\ k = 1, 2, 3$.

Linear Model : $y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$

where μ = the overall mean effect

a_i = the effect of i^{th} level of factor A

b_j = the effect of j^{th} level of factor B; $(ab)_{ij}$ is the effect of the inteaction between a_i and b_j

(e_{ijk}) = is a random error component.

$$\Sigma a_i = 0,\ \Sigma b_j = 0,\ \sum_i (ab)_{ij} = 0 = \sum_j (ab)_{ij}$$

$H_{01} : a_1 = a_2 = 0,\ H_{11}$: at least one $a_i = 0$

$H_{02} : b_1 = b_2 = 0,\ H_{12}$: at least one $b_j \neq 0$

$H_{03} : (ab)_{ij} = 0$ for all i, j; H_{13} : at least one $(ab)_{ij} \neq 0$.

$y_{1..} = 92 \quad y_{.1.} = 132 \quad y_{11.} = 56 \quad y_{12.} = 36$

$y_{2.} = 142 \quad y_{.2.} = 102 \quad y_{21.} = 76 \quad y_{22.} = 66$

$$CF = \frac{y^2...}{2 \times 2 \times 3} = \frac{234 \times 234}{12} = 4563$$

$$\sum_i \sum_j \sum_j y_{ijk}^2 = 20^2 + \ldots + 21^2 = 4886$$

$\therefore$

$$TSS = 4886 - 4563 = 323$$

$$SSA = \frac{1}{6} \sum_i y_{i..}^2 - CF = \frac{92^2 + 102^2}{2} - 4563 = 208.33$$

$$SSB = \frac{1}{6} \sum_j y_{.j.}^2 - CF = \frac{132^2 + 102^2}{2} - 4563 = 75.00$$

$$SS \text{ (sub-totals)} = \frac{1}{3} \sum_i \sum_j y_{ij.}^2 = \frac{56^2 + 36^2 + 76^2 + 66^2}{3} - 4563$$

$$= \frac{14564}{3} - 4563 = 4854.66 - 4563 = 291.66$$

$$SSAB = SS\text{(sub-totals)} - (SSA + SSB)$$

$$= 291.36 - (208.33 + 75.00) = 8.33$$

$$SSE = TSS - SS \text{ (sub-totals)} = 323 - 291.66 = 31.34$$

Hence, etc.

2^3 -FACTORIAL DESIGN

■ **Example 4** *Consider a 2^3-factorial design in two randomized blocks where the observations are as follows:*

	abc	*ab*	*ac*	*bc*	*a*	*b*	*c*	(1)
Block I :	12	13	11	14	12	11	14	13
Block II :	12	12	11	12	10	12	11	12

Analyse this data and draw your conclusions.

SOLUTION $\Sigma\Sigma x_{ij}$ = Sum of all the observations

$$= (12+13+11+14+12+11+14+13)+(12+12+11+12+10+12+11+12)$$

$$CF = \frac{192\times 192}{16} = 2304$$

$$\sum_i\sum_j x_{ij}^2 = (144+169+121+196+144+121+196+169)+(144+144+121+144+100+144+121+144) = 2322$$

$\therefore$ $$TSS = \sum_i\sum_j x_{ij}^2 - CF = 2322 - 2304 = 18$$

$$SSB = \sum_{j=1}^{2}\frac{B_j^2}{8} - CF = \frac{(100)^2+(192)^2}{8} - 2304 = 2308 - 2304 = 4$$

$$SST = \sum_{i=1}^{8}\frac{T_i^2}{2} - CF$$

$$= \tfrac{1}{2}(24^2+25^2+22^2+26^2+22^2+23^2+25^2+25^2) - CF$$

$$= \frac{4624}{2} - 2304 = 2312 - 2304 = 8$$

$\therefore$ $$SSE = TSS - (SSB + SSt) = 18 - (4+8) = 6$$

To find sum of squares for various effects, we can follow Yates method:

Treatment	I	II	Total	Column2	Column 3	Column 4	Divisor
(1)	13	12	25	47	95	192	8 × 2
a	12	10	22	48	97	−6	8 × 2
b	11	12	23	47	−1	4	8 × 2
ab	13	12	25	50	−5	6	8 × 2
c	14	11	25	−3	1	2	8 × 2
ac	11	11	22	2	3	−4	8 × 2
bc	14	12	26	−3	5	2	8 × 2
abc	12	12	24	−2	1	−4	8 × 2
Total	100	92	192	—	—		

$$SSA = \frac{(-6)^2}{16} = 2.25; \ SSB = \frac{4^2}{16} = 1.00$$

$$SAB = \frac{6^2}{10} = 2.25; \ SSC = \frac{2^2}{16} = 0.25$$

$$SAC = \frac{(-4)^2}{16} = 1.00; \ SSBC = \frac{2^2}{16} = 0.25$$

$$SSABC = \frac{(-4)^2}{16} = 1.00$$

Thus, the total *ss* due to treatments = 8.00

Now, we prepare analysis of variance table like following :

Source of Variation	*SS*	d.f.	*MSS*	F_{cal}	Conclusion
A	2.25	1	2.25	2.62	Not Significant
B	1.00	1	1.00	1.16	Not significant
AB	2.25	1	2.25	2.62	Not significant
C	0.25	1	0.25	0.29	Not significant
AC	1.00	1	1.00	1.16	Not significant
BC	0.25	1	0.25	0.29	Not significant
ABC	1.00	1	1.00	1.16	Not significant
Treatment	8.00	7	1.14	1.68	Not significant
Blocks	4.00	1	4.6	—	Significant
Error	6.00	7	0.86	4.65	—
Total	18.00	15	—	—	—

$$F_{0.05}(7, 7) = 4.53, \quad F_{0.05}(1, 7) = 5.99$$