Missing observations problem in RBD

When using RBD, sometimes one or more observations are missing in some blocks This may happens due to several reasons, such as carelessness, unavoidable damage to an experimental unit, missing error, loss of data, experimental mishaps etc. A missing observation introduces a new problem into the analysis because every treatment doesn't occur in that block.

In RBD, the statistical analysis typically relies on balanced data, meaning every treatment appears once in each block. A missing observation creates an imbalance:

- The total number of observations decreases.
- The sums of squares calculations for treatments, blocks, and error terms are affected.
- Standard ANOVA assumptions may no longer hold, leading to biased estimates of treatment effects.

The most common and widely used method for estimating missing observations in RBD is least square method. The least square method aims to minimize the error sum of squares, providing the best possible estimate of the missing value under the RBD model. It minimizes the error sum of squares (SSE) while preserving the structure of the design.

Once the missing observation is estimated, usual analysis of variance is performed just as if the estimated observation were real data, with the reduced error degrees of freedom. If one observation is missing, we reduce error d.f. by 1 and if two observations were missing, we reduce error d.f. by 2 and so on.

One observation missing problem:

Suppose the observation y_{ij} for treatment 'i' in the block 'j' is missing. Let's denote the missing observation by 'x'.

Treatment	**Block**						**Total**
	1	**2**	**…**	**j**	**…**	**b**	
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1b}	$\boldsymbol{y_{1.}}$
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2b}	$\boldsymbol{y_{2.}}$
…	…	…	…	…	…	…	**…**
i	y_{i1}	y_{i2}	…	missing obs (x)	…	y_{ib}	$\boldsymbol{y_{i.} = y'_{i.} + x}$
…	…	…	…	…	…	…	**…**
a	y_{a1}	y_{a2}	…	y_{aj}	…	y_{ab}	$\boldsymbol{y_{a.}}$
Total	$\boldsymbol{y_{.1}}$	$\boldsymbol{y_{.1}}$	**…**	$\boldsymbol{y_{.j} = y'_{.j} + x}$	**…**	$\boldsymbol{y_{.b}}$	$\boldsymbol{y_{..}}$

Here,

x = missing observation

$y'_{i.}$ = total for the treatment with one observation missing ($y_{i.} = y'_{i.} + x$)

$y'_{.j}$= total for the block with one observation missing ($y_{.j} = y'_{.j} + x$)

$y'_{..}$= grand total with one observation missing ($y_{..} = y'_{..} + x$)

We want to estimate x so that SSE is minimum

SST = $\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}$

= x^2 + terms independent of x – $\frac{(y'_{..}+x)^2}{N}$

SSA = $\frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$

= $\left(\frac{y_{1.}^2}{b} + \ldots + \frac{y_{i,}^2}{b} + \ldots + \frac{y_{a.}^2}{b}\right) - \frac{y_{..}^2}{N}$

= $\frac{(y'_{i.}+x)^2}{b}$ + terms independent of x – $\frac{(y'_{..}+x)^2}{N}$

SSB = $\frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}$

= $\left(\frac{y_{.1}^2}{a} + \ldots + \frac{y_{.j}^2}{a} + \ldots + \frac{y_{.b}^2}{a}\right) - \frac{y_{..}^2}{N}$

= $\frac{(y'_{.j}+x)^2}{a}$ + terms independent of x – $\frac{(y'_{..}+x)^2}{N}$

SSE = SST – SSA – SSB

= x^2- $\frac{(y'_{i.}+x)^2}{b}$ - $\frac{(y'_{.j}+x)^2}{a}$ - $\frac{(y'_{..}+x)^2}{N}$ + terms independent of x

Our problem is to find the value of x so that SSE is minimum.

Now,

$$\frac{\partial SSE}{\partial x} = 0$$

or, x - $\frac{y'_{i.}+x}{b}$ - $\frac{y'_{.j}+x}{a}$ + $\frac{y'_{..}+x}{ab}$ = 0

or, $\left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{ab}\right)x = \frac{y'_{i.}}{b} + \frac{y'_{.j}}{a} - \frac{y'_{..}}{ab}$

or, $\frac{(ab-a-b+1)}{ab}x = \frac{(ay'_{i.}+by'_{.j}-y'_{..})}{ab}$

or, $\hat{x} = \frac{(ay_{i.}^{/}+by_{.j}^{/}-y_{..}^{/})}{(a-1)(b-1)}$

A more simplified formula is,

$$\hat{x} = \frac{t\,T + b\,B - G}{(t-1)(b-1)}$$

where,

t = no. of treatment used

b = no. of blocks

T = total of all observations in the treatment with the missing value

B = total of all observations in the block with the missing value

G = grand total of all observed values

ANOVA Table

Source of Variation	**Degrees of freedom**	**Sum of squares**	**Mean sum of square**	**Cal F**
Treatment	a - 1	SSA*	MSA*	FA = MSA*/MSE
Block	b - 1	SSB	MSB	
Error	(a – 1) (b – 1) - 1	SSE	MSE	FB = MSB / MSE
Total	ab - 2	SST*		

Adjusted SSA (SSA*)

= Unadjusted SSA – Adjustment factor

= Unadjusted SSA - $\frac{(ay_{i.}^{/} + y_{.j}^{/} - y_{..}^{/})^2}{a(a-1)(b-1)^2}$

Missing observations problem in LSD

Missing observation in LSD

Occasionally, observations are missing in a LSD due to several reasons. Suppose an observation for ith row, jth column and receiving kth treatment is missing. Let's denote missing observation by symbol x i.e. $x = y_{ijk}$

Notations:

$y_{i..}^{/}$ = total of known observations in i^{th} row in which one observation is missing

$y'_{.j.}$ = total for the j[th] column with one observation missing

$y'_{..k}$ = total of known observations of k[th] treatment whose one observation is missing

$y'_{...}$ = Grand total with one observation missing

So,

$y_{i..} = y'_{i..} + x$, $\quad y_{.j.} = y'_{i..} + x,$ $\quad y_{..k} = y'_{..k} + x,$ $\quad y_{...} = y'_{...} + x$

Now,

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 - \frac{y_{...}^2}{N} = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 - \frac{y_{...}^2}{p^2}$$

$$= x^2 + \text{terms independent of x} - \frac{(y'_{...}+x)^2}{p^2}$$

$$SSR = \frac{1}{p}\sum_{i=1}^{p} y_{i..}^2 - \frac{y_{...}^2}{N}$$

$$= \frac{(y'_{i..}+x)^2}{p} + \text{terms independent of x} - \frac{(y'_{...}+x)^2}{p^2}$$

$$SSC = \frac{1}{p}\sum_{i=1}^{p} y_{.j.}^2 - \frac{y_{...}^2}{N}$$

$$= \frac{(y'_{.j.}+x)^2}{p} + \text{terms independent of x} - \frac{(y'_{...}+x)^2}{p^2}$$

$$SSA = \frac{1}{p}\sum_{i=1}^{p} y_{..k}^2 - \frac{y_{...}^2}{N}$$

$$= \frac{(y'_{..k}+x)^2}{p} + \text{terms independent of x} - \frac{(y'_{...}+x)^2}{p^2}$$

SSE = SST – SSR – SSC – SSA

$$= x^2 - \frac{(y'_{i..}+x)^2}{p} - \frac{(y'_{.j.}+x)^2}{p} - \frac{(y'_{..k}+x)^2}{p} + \frac{2(y'_{...}+x)^2}{p^2} + \text{terms independent of x}$$

Here,

$$\frac{\partial SSE}{\partial x} = 0$$

or, $\quad x - \frac{(y'_{i..}+x)}{p} - \frac{(y'_{.j.}+x)}{p} - \frac{(y'_{..k}+x)}{p} + \frac{2(y'_{...}+x)}{p^2} = 0$

or, $\quad x\left(1 - \frac{1}{p} - \frac{1}{p} - \frac{1}{p} + \frac{2}{p^2}\right) = \frac{p(y'_{i..}+y'_{.j.}+y'_{..k}) - 2y'_{...}}{p^2}$

or, $\quad x\left(\frac{p^2-3p+2}{p^2}\right) = \frac{p(y'_{i..}+y'_{.j.}+y'_{..k}) - 2y'_{...}}{p^2}$

or, $\quad \hat{x} = \frac{p(y'_{i..}+y'_{.j.}+y'_{..k}) - 2y'_{...}}{(p-1)(p-2)}$

More simplified formula is given by,

$$\hat{x} = \frac{k\,(R+C+T)-2G}{(k-1)(k-2)}$$

Where:

- k = number of rows (columns, treatments) in the LSD
- R = total of all observations in the row with the missing value
- C = total of all observations in the column with the missing value
- T = total of all observations in the treatment with the missing value
- G = grand total of all observed values

After estimating the missing value x, the analysis of LSD is carried out as usual expect 1 degrees of freedom is subtracted from SSE and SST degrees of freedom

The ANOVA table is as follows:

Source of variation	**Degrees of freedom**	**Sum of squares**	**Mean sum of squares**	**Cal F**
Treatment	p – 1	SSA*	MSA*	F_A = MSA*/MSE
Rows	p – 1	SSR	MSR	F_R = MSR/MSE
Column	p – 1	SSC	MSC	F_C = MSC/MSE
Error	(p – 1) (p – 2) - 1	SSE	MSC	
Total	$p^2 - 2$	SST		

SSA* = Adjusted SSA

= SSA – Adjustment Factor

$$= \text{SSA} - \left[\frac{(p-1)y'_{i..}+y'_{i..}+y'_{.j.}-y'_{...}}{(p-1)(p-2)}\right]$$