Frequently asked questions

1. What is missing plot technique?

   It may happen some time that while conducting the experiments some observations will be missed. For example, in an agricultural experiment, the seeds are sown and the yields are to be recorded after few months. It may happen that the crops of any plot is destroyed due to heavy rain, cattle grazing etc. In a clinical trail suppose the reading has to be taken after few days after giving medicines, and the patient does not turn up for reading. In such cases some of values will be missing.

In such situations one option is to estimate the missing value on the basis of the available data, replace these values in the missing places and make the data set complete. Now conduct the statistical analysis on the basis of completed data set by making necessary adjustments in the statistical tools to be applied. Such an analysis is termed as missing plot technique.

2. Estimating one missing observation in RBD:

Suppose the observation in the $(ij)^{th}$ cell is missing. Denote it by x.

The layout is given below.

			Blocks				Block Totals
Treatments	1	2		j		k	
1	y_{11}	Y_{12}	…		....	Y_{1n}	$y_{1.}$
2	y_{21}	Y_{22}.	…		....	Y_{2n}	$y_{2.}$
.	.	.	…		....	.	.
i	.	.		**Yij =x**	....	.	. T+x
.	.	.	…		....	.	.
p	y_{p1}	y_{p2}	…		....	y_{pn}	$y_{p.}$
Treatment Total	y.1	y.2	…	B + x	...	y.n	Grand Total: G + x

Where G is the total of known observations

B is the total of known observations in the j^{th} block

T is the total of known observations in the i^{th} treatment

SST = $\sum_{i=1}^{p}\sum_{j=1}^{k}(y_{ij}-\overline{y_{..}})^2$ = $\sum_{i`=1}^{p}\sum_{j`=1,(i`j`\neq ij)}^{k} y_{i`j`}^2 + x^2 - \frac{(G+x)^2_{..}}{N}$

SSTR = $\sum_{i=1}^{p} k(\overline{yi.} - \overline{y_{..}})^2$ = $\sum_{i`=1,(i`\neq i)}^{p} \frac{y_{i.`}^2}{k} + \frac{(T+x)^2}{k} - \frac{(G+x)^2}{N}$

SSB= $\sum_{j=1}^{k} p(\overline{y.j} - \overline{y_{..}})^2$ = $\sum_{j`=1(j`\neq j)}^{k} \frac{y_{.j`}^2}{p} - \frac{(B+x)_{..}^2}{N} - \frac{(G+x)^2}{N}$

And SSE =

SST-SSTR-SSB=

$\sum_{i`=1}^{p}\sum_{j`=1}^{k} y_{i`j`}^2 + x^2 - \frac{(G+x)_{..}^2}{N} - \sum_{i`=1}^{p} \frac{y_{i.`}^2}{k} + \frac{(T+x)^2}{k} - \frac{(G+x)^2}{N} -$

$\sum_{j`=1(j`\neq j)}^{k} \frac{y_{.j`}^2}{p} + \frac{(B+x)_{..}^2}{N} - \frac{(G+x)^2}{N}$

Using the principal of least squares, the estimate of x is obtained by minimising SSE.

e.i. by solving $\frac{\partial SSE}{\partial x} = 0$. Solving we have

2x- $\frac{2(T+x)}{k}$ + $\frac{2(B+x)}{p}$ + $\frac{2(G+x)}{N}$ = 0

Givers x= $\frac{pT+kB-G}{(p-1)(k-1)}$

3. Analysis of variance Table when one observation is missing in RBD

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	p-1	SSTR	MSSTR = SSTR/(p-1)	MSSTR/ MSSE
replicates	k-1	SSB	MSSB = SSB/(k-1)	MSS**B**/ MSSE
Error	(p-1) (k-1)-1	SSE	MSSE = SSE/(p-1) (k-1)	
total	N-2	TSS		

4. Estimate the missing observation , given the following information conducted in an RBD with the observation from seeding rating 100 in the 3rd replicated is missing

Table 1:

Data gives the grain in yield of rice at six seeding rates(kg.Ha)

	Seeding rates(kg.ha) Total						
Replicates	25	50	75	100	125	150	
1	5.1	5.3	5.3	5.2	4.8	5.3	31.0
2	5.4	6.0	5.7	4.8	4.8	4.5	31.2
3	5.3	4.7	5.5	??	4.4	4.9	29.8
4	4.7	4.3	4.7	4.4	4.7	4.1	26.9
Total	20.5	20.3	21.2	19.4	18.7	18.8	118.9

p

Step 1Let value from the cell (4,3) , receiving the 4th treatment in the 3rd replicate is missing. Denote it as x

Let

G is the total of known observations = 113.9

B is the total of known observations in the 3rd replicate = 24.8

T is the total of known observations in the 4th treatment = 14.4

Then the estimate of the missing observation is $x=\frac{pT+kB-G}{(p-1)(k-1)}$ = $\frac{6x14.4+4x24.8-113.9}{(6-1)(4-1)}$ = 71.7/15=4.78

5. **Write the complete ANOVA table for the data given in table 1 in Q.No.4**

The estimate of the missing value is x = 4.78. Substituting the this missing value in the missing place we have

Calculation of correction factor(CF): $\frac{y_{..}^2}{N}$ = $118.68^2/(6x4)$ = 586.87

Calculation of Total sum of squares: $\sum_{i=1}^{p}\sum_{j=1}^{k} y_{ij}^2 - \frac{y_{..}^2}{N}$ = $(5.1^2+5.4^2+\ldots+4.1^2)$-CF

=591.9184-586.87 = 5.05

Calculation of Treatment sum of squares= $\sum_{i=1}^{p} \frac{y_{i.}^2}{k} - \frac{y_{..}^2}{N}$ =

$(1/4)(20.5^2+20.3^2+21.2^2+19.19^2+18.7^2+18.8^2)$ –CF = 1.42

Calculation of Replicate sum of squares = $\sum_{i=1}^{k} \frac{y_{.j}^2}{p} - \frac{y_{..}^2}{N}$ =

$(1/6)(31.0^2+31.2^2+29.58^2+26.9^2)$ CF=1.97

Calculation of Error sum of squares

ESS = 5.02-1.2675-1.965 =1.66

Anova Table:

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	p-1=5	SSTR=1.42	0.284	2.394
replicates	k-1=3	SSB = 1.97	0.657	5.5396
Error	(p-1) (k-1)-1=14	SSE = 1.66	0.1186	
total	N-2=22	TSS= 5.05		

6. Write the procedure of ANOVA when one observation is missing

1. Express the error sum of squares as functions of missing values
2. Minimise the error sum of squares using principal of maxima and minima with respect to the missing values and obtain the estimates of these missing values
3. Replace the missing values by its estimates and complete the data set.
4. Apply analysis of variance tools.
5. The error sum of squares thus obtained is corrected but Treatment and block sum of squares are not corrected.
6. The number of degree of freedom associated with the TSS are subtracted by the number of missing values and adjusted in the error sum of squares. No change in the degree of freedom of sum of squared due to treatment or block is needed.

7. **What is the procedure adopted to estimate the missing values in ANOVA technique?**

   Using the principal of least squares, the estimate of x(the missing value)is obtained by minimising SSE. Here SSE is a function of x and is minimised w.r.t. x using maxima minima principle.

8. **Write the layout of the RBD design with 4 blocks and 5 treatments named as A,B,C,D and E and the observation from treatment A in block 3 is missing.**
9. **LAYOUT:**

Block 1	**Block 2**	**Block 3**	**Block 4**
A	**A**	**??**	**A**
B	**B**	**B**	**B**
C	**C**	**C**	**C**
D	**D**	**D**	**D**
E	**E**	**E**	**E**

10. **Following the information from an RBD design with 3 replicates and 9 treatments and observation from 3rd replicates receiving 4th treatment is missing. Complete the ANOVA table given the following information**

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	2	0.99	-	-
replicates	8	-	-	-
Error	-	-	0.574	
total	25	1585.63		

Answer:

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	2	0.99	0.495	0.8624
replicates	8	1576.03	197.0	343.61
Error	15	8.61	0.574	
total	25	1585.63		