Missing plot technique

In this session we discuss the missing plot technique. The entire topic is devided in to the following subdivisions.

1. **Meaning**
2. **Procedure**
3. Estimation the missing observation in RBD:
4. **Statistical Analysis**
5. **ANOVA Table**
6. **Example**
7. **Conclusion**

1. Meaning:

   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 cases 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. Procedure: (One observation is missing in RBD)
   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.
3. Estimating the 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= $\frac{pT+kB-G}{(p-1)(k-1)}$

4. Statistical Analysis:

The statistical Model for RBD

Y_{ij} represent the j^{th} observation taken from treatment i.

We define the model

$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, i= 1,2...p, j= 1,2 ...k

(4)

Where μ general effect

α_i – i^{th} treatment effect;

βj – j^{th} block effect;

ϵij- error term, independent and identically distributed random variables with mean 0 and variance σ2

There are two null hypothesis to be tested:

1. Related to the treatment effects

   H_A: $\alpha_1 = \alpha_2 = \ldots \alpha_p = 0$

   And the alternative hypothesis is

   H_{A1}: at least one $\alpha_i = \alpha_j$ for all I,j

2. Related to the block effects

   H_B: $\beta_1 = \beta_2 = \ldots \beta_p = 0$

   And the alternative hypothesis is

   H_{B1}: at least one$\beta_i = \beta_j$ for all I,j

   Under the null hypothesis H_A: $\alpha_1 = \alpha_2 = \ldots \alpha_p = 0$

   the ratio

$$F_A = \frac{SSTR/(p-1)}{SSE/(p-1)(k-1)-1} = \frac{MSSTR}{MSSE} \quad (1)$$

is distributed as F with p-1 and (p-1)(k-1)-1 degrees of freedom.

. Similarly if the null hypothesis H_B: $\beta_1 = \beta_2 = \ldots \beta_p = 0$

, the ratio

$$F_B = \frac{SSB/(K-1)}{SSE/(p-1)(k-1)-1} = \frac{MSSB}{MSSE} \qquad (2)$$

is distributed as F with k-1 and (p-1)(k-1)-1 degrees of freedom.

We reject H_A and conclude that there are differences in the treatment means if $F_A > F_{\alpha,,p-1,\,(p-1)(k-1)}$ Where F_0 is computed from equation 1 and Fα, p-1, (p-1)(k-1), is the table value referring to F table at α level significance corresponding to p-1and (p-1)(k-1)degrees freedom.

We reject H_B and conclude that there are differences in the group means if $F_B > F_{\alpha,,k-1,\,(p-1)(k-1)}$ Where F_B is computed from equation 2 and Fα, k-1, (p-1)(k-1), is the table value referring toF table at α level significance corresponding to k-1and (p-1)(k-1)degrees freedom.

If HB is accepted , then it indicates that the blocking is not necessary for future experimentation

5. Analysis of variance Table(ANOVA Table)

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		

6. **Example**: Consider the example discussed in discussing RBD

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

Step 1: Estimate the missing observation: The value from the cell (4,3) , receiving the 4^{th} treatment in the 3^{rd} 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 4^{th} 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

Substituting this value in the missing value in the data set. We have complete data set. Continue as in the case of RBD. I. E. After substituting the missing value in its place we have

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

Step;2: 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

Step;3: 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

Step;4: 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

Step;5: 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		

Table value for Hα is F.05, 5,15 = 2. 96

Table value for Hβ is F.05, 3,15 = 3.34

Conclusion: Since F cal(2.394) <F.05, 5,14 we do not reject Hα. Hence we conclude that all treatment means are equal

Since F cal(5.5788) >F.05, 3,14 we do reject Hβ. Hence we conclude that replicate means are significantly different ,

7. Conclusion: In this lecture we discusses the situation and meaning of missing values. We discussed the procedure when some observations are missing. We derived the estimates of missing values when one observation is missing in RBD Model.We discussed the statistical Analysis when one observation is missing under RBD model. The entire analysis we summarised in ANOVA table. We discussed the entire analysis when one observation is missing in RBD model solving a problem..