

Missing plot techniques in design of experiments

Author: Dr. A.K. Singh

Date: 2023-03-20

1. Introduction

There are many situations, some of which are enumerated below, which cause missing data or redundant data in some plots:

- i. when an animal becomes sick or dies due to which plot could be could not be ploughed and left out, causing missing plot;
- ii. when rodents destroy a plot in a field trial, causing missing data;
- iii. when a flask in laboratory breaks and observations could not be measured, causing missing data;
- iv. when there has been a significant or obvious observation-recording-error, such data should be treated as missing;
- v. due to some accidental situation, there is some loss of data of some plot, also warrants such a situation;
- vi. when errors are made in the application of the correct treatments in correct plot

2. Solution to the problem of above-mentioned missing data

Before, we begin to discuss various methods of estimating missing data, *it may be clarified beforehand, that any of the following analysis method does not, in any sense, recover the lost information due to missing data, but makes the best of what the research we have as follows:*

Method-I: Method of least squares

- i. Application of least squares analysis of the data, **in the mode of regression analysis**, solves the missing data problem, because least squares do not require symmetry in data cells, *it can be asymmetric due to missing plot.*
- ii. However, due to loss of data, the **simplicity of the analyses and the computations made are generally difficult** or unusable.

Method-II: Method of least squares estimation of missing plot

- i. Fortunately, the **missing data can be estimated by least squares and entered into the vacant cells of the table, to make the cells symmetric again**, which were destroyed due to missing data;
- ii.

Before following up with the application of usual analysis of variance, the first modification is to assume initially, that the missing plot has zero (0) observation. Then a least square analysis is done, which gives an estimation of missing observation as given in [Equation \(1\)](#) for RBD and [Equation \(4\)](#) for LSD.

- iii. Now, replacing the zero (0) used for missing plot, initially, with the estimated one, as explained above, and following up further the usual analysis, gives results correct enough for practical purposes.
- iv. However, in this particular method, there is a stringent condition that the missing data must not be due to failure of a treatment. For example, if a treatment has killed the plant(s), data of zero (0) yield should be recorded in the corresponding plot instead of treating such a plot as missing.

Method-III: Method of analysis of covariance.

- i. There is a third method too, known as **Method of analysis of covariance**. This method of covariance is in general a method of least-squares only.
- ii. Still, due to a greater length of discussion required, this method would be discussed separately, not here.
- iii. However, we shall be taking one example here for illustrations.

3. Analysis of Variance in CRD with missing plot data

- i. In a one-way classification of complete randomization design, the effect of missing values is merely to reduce the sample sizes in the affected classes.
- ii. The analysis of variance is **very easily handled correctly by the methods for one-way classifications with unequal numbers of observation per cell.**

4. Analysis of Variance in randomized block design (RBD) with missing plot data

Contrary to the missing plot technique in CRD, the RBD is a balanced design, **in which levels of all treatments as well as those of blocks should be equal.** Hence, it is important to estimate the missing value. Since [Method-I](#) and [Method-III](#) are both a little cumbersome, we shall be resorting to [Method-II](#) here.

In randomized blocks, a single missing value is estimated by (\hat{X}) in the [Equation \(1\)](#), using the least square method,

$$\hat{X} = \frac{tT + rB + G}{(t - 1)(r - 1)} \quad (1)$$

where,

t = number of treatments,

r = number of blocks,

T = treatment total considering missing plot values as zero,

B = block total considering missing plot values as zero,

G = grand total of all observations considering missing plot values as zero.

Thus, the step wise procedure goes like as follows:

- i. Assume initially that the missing plot has zero (0) value.
- ii. Then, find out the totals corresponding to treatments (T) and blocks (B), apart from finding the grand total (G).
- iii. Estimate the missing plot value (X) as per the formula given in [Equation \(1\)](#).
- iv. Now, replace the missing plot value zero with that of \hat{X} .
- v. And then, carry out the usual analysis of variance (ANOVA), with some modifications, as enunciated below:
 - a. Degree of freedoms for total and of error are each reduced by 1, due to one estimated value (\hat{X}) of missing plot.
 - b. The Treatments mean square in the analysis of variance is **slightly inflated**. *The correction for this upward bias is to subtract from the treatment mean square the value (\hat{Y}) as obtained in [Equation \(2\)](#).*

$$\hat{Y} = \frac{[B - (t - 1)X]^2}{t(t - 1)^2} \quad (2)$$

- c. For multiple comparison etc., the usual standard error (i.e., s) is modified to, as in [Equation \(3\)](#).

$$s = \sqrt{s^2 \left[\frac{2}{r} + \frac{t}{r(r - 1)(t - 1)} \right]} \quad (3)$$

where, s^2 represents the corrected error mean square after dividing error sum of squares by the corrected error degrees of freedom, which is reduced by 1 due to estimation of one missing value.

Example 1. ANOVA with one missing observation in RBD setup

The present example data, vide [Table 1](#), “[Yield of wheat strain in RBD with one missing value](#)”, shows the yield output of an experiment with four strains of *Gallipoli* wheat, in which it is assumed that the yield for strain 4 (D) in block 1 is missing. Estimate this missing value and carry out usual ANOVA afterwards, i.e., carry out following steps of computations:

- i. Estimate the missing value.
- ii. Carry out ANOVA considering the effect of missing value in ANOVA.

Table 1. Yield of wheat strain in RBD with one missing value

| Strain | Block1 | Block2 | Block3 | Block4 | Block5 |
|--------|--------|--------|--------|--------|--------|
| A | 32.3 | 34 | 34.3 | 35 | 36.5 |
| B | 33.3 | 33 | 36.3 | 36.8 | 34.5 |
| C | 30.8 | 34.3 | 35.3 | 32.3 | 35.8 |
| D | NA | 26 | 29.8 | 28 | 28.8 |

Solution:

a. Applying method of estimating one missing data vide Equation (1):

Beginning and ending 6 rows of dataset for brevity

```
data.yield = read.csv("wheat-strain-yield-in-rbd.csv", header=TRUE);
data.yield.long = reshape(data.yield[,2:6], v.names = "Yield",
varying=list(names(data.yield[,c(2:6)])), times=c("Block1", "Block2", "Block3", "Block4",
"Block5"), timevar = "Block",
idvar="Strain", ids=row.names(data.yield[,1:4]), direction="long");
data.yield.long=data.yield.long[,c(1,3,2)];
save.image("wheat-strain-yield-in-rbd.RData");
#Beginning 6 rows
head(data.yield.long);
```

```
      Block Strain Yield
1.Block1 Block1      1 32.3
2.Block1 Block1      2 33.3
3.Block1 Block1      3 30.8
4.Block1 Block1      4  NA
1.Block2 Block2      1 34.0
2.Block2 Block2      2 33.0
```

```
#Ending 6 rows
tail(data.yield.long);
```

```
      Block Strain Yield
3.Block4 Block4      3 32.3
4.Block4 Block4      4 28.0
1.Block5 Block5      1 36.5
2.Block5 Block5      2 34.5
3.Block5 Block5      3 35.8
4.Block5 Block5      4 28.8
```

Step-1: In the above data.frame we now assume initially that the missing plot has zero (0) value in place of NA, as follows, in a new variable 'Yield0':

```
data.yield.long$Yield0=with(data.yield.long, replace(Yield, which(is.na(Yield)),0))
```

Step-2: Finding out the totals corresponding to treatments (*T*) and blocks (*B*), apart from finding the grand total (*G*) as follows, with missing plot having zero(0) value, as follows:

```
#Computing indices of the cell with missing value
Bmi=with(data.yield.long, Block[which(is.na(Yield))]) #Block missing index
Tmi=with(data.yield.long, Strain[which(is.na(Yield))]) #Treat missing index
#Determining some constants
n=nrow(data.yield.long);
r=with(data.yield.long,length(unique(Block)));
k=with(data.yield.long,length(unique(Strain)));
#Computing Grand Total sum with missing data
G=with(data.yield.long, sum(Yield0));
#Computing Block sum with missing data
BY=with(data.yield.long, tapply(Yield0,list(Block),sum));
B=BY[Bmi]; #Block column total with missing data
#Computing Treatment sum with missing data
TY=with(data.yield.long, tapply(Yield0,list(Strain),sum));
T=TY[Tmi]; #Treatment row total with missing data
#Treatment total (T), Block total (B), Grand total (G) with zero(0) missing value
c(T, B, G);
```

```
      4 Block1
112.6  96.4 627.1
```

Step-3: Estimating the missing plot value (*X*) as per the formula given in [Equation \(1\)](#), as follows:

```
X=(k*T+r*B-G)/((k-1)*(r-1)); X
```

```
      4
25.44167
```

Step-4: In the above data.frame we now replace the missing plot value 'NA' with that of the estimated \hat{X} , as follows in a new variable 'YieldX':

```
data.yield.long$YieldX=with(data.yield.long, replace(Yield, which(is.na(Yield)),X));
#Beginning 6 rows, containing replacement with estimated value (X)
head(data.yield.long);
```

```
      Block Strain Yield Yield0 YieldX
1.Block1 Block1     1  32.3   32.3 32.30000
2.Block1 Block1     2  33.3   33.3 33.30000
```

| | | | | | |
|----------|--------|---|------|------|----------|
| 3.Block1 | Block1 | 3 | 30.8 | 30.8 | 30.80000 |
| 4.Block1 | Block1 | 4 | NA | 0.0 | 25.44167 |
| 1.Block2 | Block2 | 1 | 34.0 | 34.0 | 34.00000 |
| 2.Block2 | Block2 | 2 | 33.0 | 33.0 | 33.00000 |

b. ANOVA procedure after estimating missing data (X)

Step-5: Now, carrying out the usual analysis of variance (ANOVA), with some modifications, as enunciated below:

- i. Degree of freedoms for total and of error are each reduced by 1, due to one estimated value (\hat{X}) of missing plot.
- ii. The Treatments mean square in the analysis of variance is **slightly inflated**. *The correction for this **upward bias** is to subtract from the treatment mean square the value (\hat{Y}) as obtained in [Equation \(2\)](#).*
- iii. The F-ratio and P-values of treatment are shown in two columns, the usual one and the corrected one due to estimation.

```

#Declaring factors
data.yield.long$Block=factor(data.yield.long$Block);
data.yield.long$Strain=factor(data.yield.long$Strain);
#Computing Degrees of Freedom (Unadjusted)
Sdf=n-1;
Bdf=r-1;
Tdf=k-1;
Edf=Sdf-Bdf-Tdf;
#Computing Total sums of squares
GY=with(data.yield.long, sum(YieldX));
CF.YY=with(data.yield.long, (GY^2/n));
SYY=with(data.yield.long, sum(YieldX^2)-CF.YY);
#Computing Block sums of squares
BY=with(data.yield.long, tapply(YieldX,list(Block),sum));
BYY=with(data.yield.long, sum(BY^2/k)-CF.YY);
#Computing Treatment sums of squares
TY=with(data.yield.long, tapply(YieldX,list(Strain),sum));
TYY=with(data.yield.long, sum(TY^2/r)-CF.YY);
#Computing Error sums of squares
EYY=SYY-BYY-TYY;
#Computing Mean Squares
SMS=SYY/Sdf;
BMS=BYY/Bdf;
TMS=TYY/Tdf; #uncorrected
EMS=EYY/Edf; #uncorrected
#Adjustments/correction due to missing data (X)
Sdf.corrected=Sdf-1; #Total df correction
Edf.corrected=Edf-1; #Error df correction
TMS.correction=(B-(k-1)*X)^2/(k*(k-1)^2);
TMS.corrected=TMS-TMS.correction;
EMS.corrected=EYY/Edf.corrected;
#F-ratio of Blocks and treatment
F.B=BMS/EMS;
F.T=TMS/EMS;
F.T.corrected=TMS.corrected/EMS.corrected;
#P-value for Block and treatment
Pv.F.B=1-pf(F.B,Bdf,Edf);
Pv.F.T=1-pf(F.T,Tdf,Edf);
Pv.F.T.corrected=1-pf(F.T.corrected,Tdf,Edf.corrected);
#Preparing ANOVA structure
SV=c("Total","Block","Treat","Error");
DF=c(Sdf.corrected,Bdf,Tdf,Edf.corrected);
SSY=c(SYY,BYY,TYY,EYY);
MSY=c(SMS,BMS,TMS,EMS);
MSY.corrected=c(SMS,BMS,TMS.corrected,EMS.corrected);
F.ratio=c(NA,F.B,F.T,NA);
F.ratio.corrected=c(NA,F.B,F.T.corrected,NA);
Pv=c(NA,Pv.F.B,Pv.F.T,NA);
Pv.corrected=c(NA,Pv.F.B,Pv.F.T.corrected,NA);
#ANOVA in RBD with estimate of one missing data
ANOVA.RBD=data.frame(SV,DF,SSY,MSY,MSY.corr=MSY.corrected,F=F.ratio,F.corr=F.ratio.corrected,Pv,Pv.corr=Pv.corrected);
ascii::ascii(ANOVA.RBD,include.rownames=TRUE,include.colnames=TRUE,header=TRUE);

```

| | SV | DF | SSY | MSY | MSY.corr | F | F.corr | Pv | Pv.corr |
|---|-------|-------|--------|-------|----------|-------|--------|------|---------|
| 1 | Total | 18.00 | 223.48 | 11.76 | 11.76 | | | | |
| 2 | Block | 4.00 | 35.20 | 8.80 | 8.80 | 6.09 | 6.09 | 0.01 | 0.01 |
| 3 | Treat | 3.00 | 170.94 | 56.98 | 45.79 | 39.46 | 29.06 | 0.00 | 0.00 |
| 4 | Error | 11.00 | 17.33 | 1.44 | 1.58 | | | | |

c. Multiple comparison of pairs of means

i. Step-6: Computing Treatment means

```
Strain.mean.corr=with(data.yield.long, tapply(YieldX, list(Strain),mean));
Strain.mean.corr
```

```
      1      2      3      4
34.42000 34.78000 33.70000 27.60833
```

ii. Step-7: Computing Difference of Treatment means

```
Strain.mean.corr.diff=with(data.yield.long, outer(Strain.mean.corr,Strain.mean.corr,"-"))
Strain.mean.corr.diff
```

```
      1      2      3      4
1  0.000000 -0.360000  0.720000  6.811667
2  0.360000  0.000000  1.080000  7.171667
3 -0.720000 -1.080000  0.000000  6.091667
4 -6.811667 -7.171667 -6.091667  0.000000
```

iii. Step-8: Computing Corrected Standard Errors for difference among means

```
EMS.diff.means=EMS.corrected*(2/r+k/(r*(r-1)*(k-1)));
se.diff.means=sqrt(EMS.diff.means);
c(EMS.diff.means," ", se.diff.means);
```

```
[1] "0.735210353535447" "0.857444081870909"
```

iv. Step-9: Computing Critical Difference

```
alpha=0.05;
CD=abs(qt(1-alpha/2,Edf.corrected))*se.diff.means
CD
```

```
[1] 1.887222
```

v.

Step-10: Comparison among pairs of treatment means and CD at 5% level of significance

```
Multi.comp.diff.Strain.means.corr=(Strain.mean.corr.diff > CD)
Multi.comp.diff.Strain.means.corr
```

```
      1      2      3      4
1 FALSE FALSE FALSE TRUE
2 FALSE FALSE FALSE TRUE
3 FALSE FALSE FALSE TRUE
4 FALSE FALSE FALSE FALSE
```

From above results of comparison of differences of treatment means with critical difference at 5% level of significance, it may be concluded that treatment 4 (i.e. Strain 'D') has significantly different effect on Yield compared to those of treatments 1,2 and 3 (i.e. Strains A, B, C). On the contrary one can also interpret that since yield of treatment D is minimum, all the Strains (A, B and C) are having significantly more yield than that of Strain D, while remaining Strains are at par at 5% level of significance.

d. Saving workspace image

```
save.image("wheat-strain-yield.RData");
```

5. Analysis of Variance in latin square design (LSD) with missing plot data

Contrary to the missing plot technique in CRD, the LSD is also a balanced design, **in which levels of all treatments as well as those of the row and column blocks should have all equal number observation**. Note, this design is conducted in two dimension but its computations are actually in 3 dimensions, namely, treatment, row block and column block Hence, it is important to estimate the missing value. Since [Method-I](#) and [Method-III](#) are both a little cumbersome, we shall be resorting, here too, to [Method-II](#).

In latin square design blocks. a single missing value is estimated by (\hat{X}^*) in the [Equation \(4\)](#), using the least square method,

$$\hat{X}^* = \frac{t(R + C + T) - 2G}{(t - 1)(t - 2)} \quad (4)$$

where,

t = number of treatments, or rows or columns

T = treatment total considering missing plot values as zero,

R = row total considering missing plot values as zero,

C = column total considering missing plot values as zero,

G = grand total of all observations considering missing plot values as zero.

Thus, the step wise procedure goes like as follows:

- i. Assume initially that the missing plot has zero (0) value.
- ii. Then, find out the totals corresponding to treatments (T), row blocks (R), column blocks (C), apart from finding the grand total (G).
- iii. Estimate the missing plot value (X) as per the formula given in [Equation \(4\)](#).
- iv. Now, replace the missing plot value zero with that of \hat{X}^* .
- v. And then, carry out the usual analysis of variance (ANOVA), with some modifications, as enunciated below:
 - a. Degree of freedoms for total and of error are each reduced by 1, due to one estimated value (\hat{X}^*) of missing plot.
 - b. The Treatments mean square in the analysis of variance is **slightly inflated**. *The correction for this **upward bias** is to subtract from the treatment mean square the value (\hat{Y}) as obtained in [Equation \(5\)](#).*

$$\hat{Y}^* = \frac{[G - R - C - (t - 1)T]^2}{(t - 1)^3(t - 2)^2} \quad (5)$$

- c. For multiple comparison etc., the usual standard error (i.e., s) is modified to, as in [Equation \(6\)](#).

$$s^* = \sqrt{s^{*2} \left[\frac{2}{t} + \frac{1}{(t - 1)(t - 2)} \right]} \quad (6)$$

where, s^{*2} represents the corrected error mean square after dividing error sum of squares by the corrected error degrees of freedom, which is reduced by 1 due to estimation of one missing value.

Example 2. ANOVA with one missing observation in LSD setup

The present example data, vide [Table 2](#), “**Milk Yield of Cows in LSD with one missing value**”, pertains to experiments, presented in Statistical Methods (6th ed.) by Snedecor and Cochran, affecting the milk yield of dairy cows. To assess the great variations among cows requires large numbers of animals for evaluating moderate differences. Efforts to apply several treatments successively to the same cow are complicated by the decreasing milk flow, by the shapes of the lactation curves. by carry over effects, and by presumed correlation among the errors, ϵ_{ij} . The effort was made to control these difficulties by the use of several pairs of orthogonal latin squares, the columns representing cows, the rows successive periods during lactation, the treatments being A = roughage, B = limited grain, C = full grain. However, for this experiment, at the moment, data of a single latin square is only presented below for brevity and also no effort is being made to deal with the carryover effects. The entries are pounds of milk for a 6-week period.

Now suppose that the milk yield 608 pounds for Cow-I and Period-I was missing. Then,

- i. Estimate the missing value.
- ii. Carry out ANOVA considering the effect of missing value in ANOVA.

Table 2. Milk Yield of Cows in LSD with one missing value

| Period | Cow-I-tr | Cow-I | Cow-II-tr | Cow-II | Cow-III-tr | Cow-III |
|--------|----------|-------|-----------|--------|------------|---------|
| I | A | 608 | B | 885 | C | 940 |
| II | B | 715 | C | 1087 | A | 766 |
| III | C | 844 | A | 711 | B | 832 |

Solution:

a. Applying method of estimating one missing data vide Equation (4):

Beginning and ending 6 rows of dataset for brevity

```

milk.yield = read.csv("cows-milk-yield-in-lsd.csv", header=TRUE);
milk.yield.long.yld = reshape(milk.yield[,c(3,5,7)], v.names = "Milk.Yield",
varying=list(names(milk.yield[,c(3,5,7)])), times=c("Cow.I", "Cou.II", "Cow.III"), timevar =
"Col.block",
idvar="Period", ids=row.names(milk.yield), direction="long");
milk.yield.long.tr = reshape(milk.yield[,c(2,4,6)], v.names = "Treat",
varying=list(names(milk.yield[,c(2,4,6)])), times=c("Cow.I.tr", "Cou.II.tr",
"Cow.III.tr"), timevar = "Col.block",
idvar="Period", ids=row.names(milk.yield), direction="long");
milk.yield.long=data.frame(milk.yield.long.yld,Treat=milk.yield.long.tr[,c(2)]);
save.image("cows-milk-yield-in-lsd.RData");
#Beginning 6 rows
head(milk.yield.long);

```

```

          Col.block Milk.Yield Period Treat
1.Cow.I      Cow.I      608      1      A
2.Cow.I      Cow.I      715      2      B
3.Cow.I      Cow.I      844      3      C
1.Cou.II     Cou.II      885      1      B
2.Cou.II     Cou.II     1087      2      C
3.Cou.II     Cou.II      711      3      A

```

```

#Ending 6 rows
tail(milk.yield.long);

```

```

          Col.block Milk.Yield Period Treat
1.Cou.II     Cou.II      885      1      B
2.Cou.II     Cou.II     1087      2      C
3.Cou.II     Cou.II      711      3      A
1.Cow.III    Cow.III      940      1      C
2.Cow.III    Cow.III      766      2      A
3.Cow.III    Cow.III      832      3      B

```

Step-1: In the above data.frame we now assume initially that the missing plot has zero (0) value in place of 608, as follows, in a new variable '**Yield0**':

```
milk.yield.long$Milk.Yield0=milk.yield.long$Milk.Yield;
milk.yield.long[milk.yield.long$Milk.Yield0==608,"Milk.Yield0"]=0;
```

Step-2: Finding out the totals corresponding to treatments (*T*), row blocks (*R*) and column blocks (*C*), apart from finding the grand total (*G*) as follows, with missing plot having zero(0) value, as follows:

```
#Computing indices of the cell with missing value
Rmi=with(milk.yield.long, Period[which(milk.yield.long$Milk.Yield0==0)]); #Row Block
missing index
Cmi=with(milk.yield.long, Col.block[which(milk.yield.long$Milk.Yield0==0)]); #Column Block
missing index
Tmi=with(milk.yield.long, Treat[which(milk.yield.long$Milk.Yield0==0)]); #Treat missing
intex
#Determining some constants
n=nrow(milk.yield.long);
t=with(milk.yield.long,length(unique(Treat)));
#Computing Grand Total sum with missing value
G=with(milk.yield.long, sum(Milk.Yield0));
#Computing Row-Block sum with missing value
RY=with(milk.yield.long, tapply(Milk.Yield0,list(Period),sum));
R=RY[Rmi]; #Row-Block total with missing milk
#Computing Column-Block sum with missing value
CY=with(milk.yield.long, tapply(Milk.Yield0,list(Col.block),sum));
C=CY[Cmi]; #Column-Block total with missing milk
#Computing Treatment sum with missing value
TY=with(milk.yield.long, tapply(Milk.Yield0,list(Treat),sum));
T=TY[Tmi];
#Row-block total (B), Column-block total, Treatment total (T), and Grand total (G) with
zero(0) missing value
c(R, C, T, G);
```

```
1 Cow.I      A
1825 1559 1477 6780
```

Step-3: Estimating the missing plot value (\hat{X}^*) as per the formula given in [Equation \(4\)](#), as follows:

```
X.=(t*(R+C+T)-2*G)/((t-1)*(t-2)); X.
```

```
1
511.5
```

Step-4: In the above data.frame we now replace the missing plot value '0' with that of the estimated \hat{X}^* , as follows in a new variable '**YieldX**.':

```

milk.yield.long$Milk.YieldX.=with(milk.yield.long, replace(Milk.Yield,
which(milk.yield.long$Milk.Yield0==0),X.));
#Beginning 6 rows, containing replacement with estimated value (X)
head(milk.yield.long);

```

| | Col.block | Milk.Yield | Period | Treat | Milk.Yield0 | Milk.YieldX. |
|----------|-----------|------------|--------|-------|-------------|--------------|
| 1.Cow.I | Cow.I | 608 | 1 | A | 0 | 511.5 |
| 2.Cow.I | Cow.I | 715 | 2 | B | 715 | 715.0 |
| 3.Cow.I | Cow.I | 844 | 3 | C | 844 | 844.0 |
| 1.Cou.II | Cou.II | 885 | 1 | B | 885 | 885.0 |
| 2.Cou.II | Cou.II | 1087 | 2 | C | 1087 | 1087.0 |
| 3.Cou.II | Cou.II | 711 | 3 | A | 711 | 711.0 |

b. ANOVA procedure after estimating missing data (\hat{X}^*)

Step-5: Now, carrying out the usual analysis of variance (ANOVA), with some modifications, as enunciated below:

- i. Degree of freedoms for total and of error are each reduced by 1, due to one estimated value (\hat{X}^*) of missing plot.
- ii. The Treatments mean square in the analysis of variance is **slightly inflated**. *The correction for this **upward bias** is to subtract from the treatment mean square the value (\hat{Y}^*) as obtained in [Equation \(5\)](#).*
- iii. The F-ratio and P-values of treatment are shown in two columns, the usual one and the corrected one due to estimation.

```

#Declaring factors
milk.yield.long$Col.block=factor(milk.yield.long$Col.block);
milk.yield.long$Period=factor(milk.yield.long$Period);
milk.yield.long$Treat=factor(milk.yield.long$Treat);
#Computing Degrees of Freedom (Unadjusted)
Sdf=n-1;
Rdf=t-1;
Cdf=t-1;
Tdf=t-1;
Edf=Sdf-Rdf-Rdf-Tdf;
#Computing Total sums of squares
GY=with(milk.yield.long, sum(Milk.YieldX.));
CF.YY=with(milk.yield.long, (GY^2/n));
SYY=with(milk.yield.long, sum(Milk.YieldX.^2)-CF.YY);
#Computing Row-Block sums of squares
RY=with(milk.yield.long, tapply(Milk.YieldX.,list(Period),sum));
RYY=with(milk.yield.long, sum(RY^2/t)-CF.YY);
#Computing Column-Block sums of squares
CY=with(milk.yield.long, tapply(Milk.YieldX.,list(Col.block),sum));
CYY=with(milk.yield.long, sum(CY^2/t)-CF.YY);
#Computing Treatment sums of squares
TY=with(milk.yield.long, tapply(Milk.YieldX.,list(Treat),sum));
TYY=with(milk.yield.long, sum(TY^2/t)-CF.YY);
#Computing Error sums of squares
EYY=SYY-RYY-CYY-TYY;
#Computing Mean Squares
SMS=SYY/Sdf;
RMS=RYY/Rdf;
CMS=CYY/Cdf;
TMS=TYY/Tdf; #uncorrected
EMS=EYY/Edf; #uncorrected
#Adjustments/correction due to missing data (X)
Sdf.corrected=Sdf-1; #Total df correction
Edf.corrected=Edf-1; #Error df correction
TMS.correction=(G-R-C-(t-1)*T)^2/((t-1)^3*(t-2)^2);
TMS.corrected=TMS-TMS.correction;
EMS.corrected=EYY/Edf.corrected;
#F-ratio of Row, Column Blocks and Treatment
F.R=RMS/EMS;
F.C=CMS/EMS;
F.T=TMS/EMS;
F.T.corrected=TMS.corrected/EMS.corrected;
#P-value for Row, Column Blocks and Treatment
Pv.F.R=1-pf(F.R,Rdf,Edf);
Pv.F.C=1-pf(F.C,Cdf,Edf);
Pv.F.T=1-pf(F.T,Tdf,Edf);
Pv.F.T.corrected=1-pf(F.T.corrected,Tdf,Edf.corrected);
#Preparing ANOVA structure
SV=c("Total","Period","Cow","Treat","Error");
DF=c(Sdf.corrected,Rdf,Cdf,Tdf,Edf.corrected);
SSY=c(SYY,RYY,CYY,TYY,EYY);
MSY=c(SMS,RMS,CMS,TMS,EMS);
MSY.corrected=c(SMS,RMS,CMS,TMS.corrected,EMS.corrected);
F.ratio=c(NA,F.R,F.C,F.T,NA);
F.ratio.corrected=c(NA,F.R,F.C,F.T.corrected,NA);
Pv=c(NA,Pv.F.R,Pv.F.C,Pv.F.T,NA)
Pv.corrected=c(NA,Pv.F.R,Pv.F.C,Pv.F.T.corrected,NA)
#ANOVA in LSD with estimate of one missing data
ANOVA.LSD=data.frame(SV,DF,SSY,MSY,MSY.corrected,F=F.ratio,F.corr=F.ratio.corr

```

| | SV | DF | SSY | MSY | MSY.corr | F | F.corr | Pv | Pv.corr |
|---|--------|------|-----------|----------|----------|-------|--------|------|---------|
| 1 | Total | 7.00 | 210758.00 | 26344.75 | 26344.75 | | | | |
| 2 | Period | 2.00 | 9878.17 | 4939.08 | 4939.08 | 3.56 | 3.56 | 0.22 | 0.22 |
| 3 | Cow | 2.00 | 68304.17 | 34152.08 | 34152.08 | 24.63 | 24.63 | 0.04 | 0.04 |
| 4 | Treat | 2.00 | 129802.17 | 64901.08 | 40480.58 | 46.80 | 14.60 | 0.02 | 0.18 |
| 5 | Error | 1.00 | 2773.50 | 1386.75 | 2773.50 | | | | |

c. Multiple comparison of pairs of means

i. Step-6: Computing Treatment means

```
Treat.mean.corr=with(milk.yield.long, tapply(Milk.YieldX., list(Treat),mean));
Treat.mean.corr;
```

```
      A      B      C
662.8333 810.6667 957.0000
```

ii. Step-7: Computing Difference of Treatment means

```
Treat.mean.corr.diff=with(milk.yield.long, outer(Treat.mean.corr,Treat.mean.corr,"-"));
Treat.mean.corr.diff;
```

```
      A      B      C
A  0.0000 -147.8333 -294.1667
B 147.8333  0.0000 -146.3333
C 294.1667 146.3333  0.0000
```

iii. Step-8: Computing Corrected Standard Errors for difference among means

```
EMS.diff.means=EMS.corrected*(2/t+1/((t-1)*(t-2)));
se.diff.means=sqrt(EMS.diff.means);
c(EMS.diff.means,",", se.diff.means);
```

```
[1] "3235.75000000109" "," "56.8836531878982"
```

iv. Step-9: Computing Critical Difference

```
alpha=0.05;
CD=abs(qt(1-alpha/2,Edf.corrected))*se.diff.means
CD
```

```
[1] 722.7753
```

v. Step-10: Comparison among pairs of treatment means and CD at 5% level of significance

```
Multi.comp.diff.Treat.means.corr=(Treat.mean.corr.diff > CD);  
Multi.comp.diff.Treat.means.corr;
```

```
      A      B      C  
A FALSE FALSE FALSE  
B FALSE FALSE FALSE  
C FALSE FALSE FALSE
```

From above results of comparison of differences of treatment means with critical difference at 5% level of significance, it may be concluded that treatment 4 (i.e. Strain 'D') has significantly different effect on Yield compared to those of treatments 1,2 and 3 (i.e. Strains A, B, C). On the contrary one can also interpret that since yield of treatment D is minimum, all the Strains (A, B and C) are having significantly more yield than that of Strain D, while remaining Strains are at par at 5% level of significance.

d. Saving workspace image

```
save.image("cows-milk-yield-in-lsd.RData");
```