PRESIDENCY UNIVERSITY
KOLKATA

COMPARISON OF DIFFERENT TESTING PROCEDURES FOR ONE MISSING DATA IN RBD

Akash Das
Niranjan Dey

Project Under Prof. Saurav De

Acknowledgment :

We would like to thank and express our heartiest gratitude to our respected **Prof. Saurav De** sir for providing us with this interesting project and help us to learn some interesting result and facts regarding the topic. We shall remain ever grateful to him for pushing our limit into real life working field.

Abstract

One of the most popular design for analyzing the effects of different treatments among different groups of population is Rectangular Block Design (RBD). In practice we may come across situations where one observation is missing due to some reason. This leads to a very big problem in RBD set up, as we can't just remove the missing data from our design, unlike CRD. Thus various missing plot techniques can be performed to construct the concerned tests which reflects an approximation of the true nature the population characteristics, we are interested in. This project is mainly focused on investigating numerically how much accurate these testing procedures are. We use simulation to visualize and demonstrate the relation and accuracy of different testing procedure empirically.

1 Introduction

The data we will be working with, looks like the following,

Blocks	Treatments					
	T_1	T_2	$\dots$	T_j	$\dots$	T_v
B_1	y_{11}	y_{12}	$\dots$	y_{1j}	$\dots$	y_{1v}
B_2	y_{21}	y_{22}	$\dots$	y_{2j}	$\dots$	y_{2v}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_i	y_{i1}	y_{i2}	$\dots$	$*$	$\dots$	y_{iv}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_b	y_{b1}	y_{b2}	$\dots$	y_{bj}	$\dots$	y_{bv}

Table 1: General form of RBD with one missing data at $(i,j)^{\text{th}}$cell

We have v treatments and b blocks with one observation per cell. One data at $(i,j)^{\text{th}}$cell is missing. In case of CRD we may ignore the data which is missing but, RBD doesn't allow us to do so. So, to perform the testing regarding the treatment effects we may adopt some different procedures.

- For the time being let us consider that the observation for treatment-1 of block-1 i.e. y_{11}, is missing .

Note: Is this consideration valid?

Later we will consider different values of block effects and treatment effects randomly. So effectively the missing observation can be in any place of the table.

2 Approximate Testing Procedure :

The method of analysis of experiments with missing observations by estimating the missing observations is due to Yates (1937).

We calulate,

$$B_1' = \text{total of all available observations for block } 1 = \sum_{j(\neq 1)} y_{1j}$$
$$T_1' = \text{total of all available observations for treatment } 1 = \sum_{i(\neq 1)} y_{i1}$$

$$G' = \text{total of all available } (bv-1) \text{ observations for } = \sum_{\substack{i \\ (i,j)\neq(1,1)}}\sum_{j} y_{ij}$$

Let the missing observation, $y_{11} = x$.
So, different sum of squares are as follows,

$$SS_{BL} = \frac{(B_1'+x)^2+\sum_2^b B_i^2}{v} - \frac{(G'+x)^2}{bv}$$

$$SS_{TR} = \frac{(T'+x)^2+\sum_2^v T_i^2}{b} - \frac{(G'+x)^2}{bv}$$

$$TSS = \sum_{\substack{i \\ (i,j)\neq(1,1)}}\sum_{j} y_{ij}^2 + x^2 - \frac{(G'+x)^2}{bv}$$

$$\begin{aligned} SSE &= TSS - SS_{BL} - SS_{TR} \\ &= x^2 + \frac{(G'+x)^2}{bv} - \frac{(B_1'+x)^2}{v} - \frac{(T'+x)^2}{b} + \text{ terms not involving } x \end{aligned}$$

Minimizing this SSE with respect to x we obtain,

$$\frac{d(SSE)}{dx} = 2x + \frac{2(G'+x)}{bv} - \frac{2(B'+x)}{v} - \frac{2(T'+x)}{b} = 0$$

$$\implies \hat{x} = \frac{bB_1'+vT_1'-G'}{(b-1)(v-1)}$$

$\hat{x}$ is the LSE of the yield of the missing plot.

Next we imputed this value in the table of observations. The marginal means of the augmented table gives the block and the treatment means and comparison of treatment effects are obtained directly from the same comparison of treatment means from the augmented table.

We want to test,

$$\begin{aligned} &H_0 : \tau_1 = \tau_2 = ... = \tau_v = 0 \\ &vs \\ &H_1 : \text{ at least one of the values are different from } 0 \end{aligned}$$

Using $\hat{x}$ we calculate $SS_{BL}, SS_{TR,}SSE$ and TSS. Same as we do in ordinary RBD.

Sources	$d.f$	SS	$MS = SS/d.f$
Blocks	$b-1$	$\frac{(B_1'+\hat{x})^2+\sum_{i=2}^{b} B_i^2}{v} - \frac{(G'+\hat{x})^2}{bv}$	MS_{BL}
Treatments	$v-1$	$\frac{(T_1'+\hat{x})^2+\sum_{2}^{v} T_i^2}{b} - \frac{(G'+\hat{x})^2}{bv}$	MS_{TR}
Error	$(b-1)(v-1)-1$	$SSE(\hat{x}) = TSS - SS_{BL} - SS_{TR}$	MSE
TSS	$bv-2$	$\sum_{i}\sum_{j}{}_{(i,j)\neq(1,1)} \; y_{ij}^2 + \hat{x}^2 - \frac{(G'+\hat{x})^2}{bv}$	—

Table 2: Augmented ANOVA table for Approximate testing procedure-1

An approximate test statistic of the hypothesis is, $\frac{MS_{TR}(\hat{x})}{MSE} = \frac{SS_{H_0}/(v-1)}{SSE(\hat{x})/(bv-v-b)}$, from the augmented table which follows approximately $F_{(v-1),(b-1)(v-1)-1}$ under H_0.

Here, 1 error d.f is lost due to estimation of y_{11} by $\hat{x}$. In general, if L observations are missing and are estimated, then d.f s of SSE and TSS will be $(b-1)(v-1)-L$ and $bv-L$ respectively.

In the general case when the observation y_{kl} corresponding to k^{th} block and l^{th} treatment is missing , the best estimate of yield for this plot is ,

$$\hat{x} = \frac{bB_k' + vT_l' - G'}{(b-1)(v-1)}$$

where B_k' is the sum of all available observations for the k^{th} block and T_l' is the sum of all available observations for the l^{th} treatment and G' is the total of all available observations.

Biasedness of this test : This test is biased in the sense that expectation of the treatment MS is greater than the expectation of the error MS even under null hypothesis.

explanation :

We know that $E(MSE) = \sigma^2$ and when no data is missing, $E(MS_{TR}) = \sigma^2 + b\sigma_t^2$ where, $\sigma_t^2 = \frac{1}{v-1}\sum_{j=1}^{v}\tau_j^2$

So, under null hypothesis, $E(MS_{TR}) = \sigma^2$, as $\sigma_t^2 = \frac{1}{v-1}\sum_{j=1}^{v}(0)^2 = 0$

Hence it is the valid error for testing the null hypothesis, $H_0 : \tau_1 = \tau_2 = ... = \tau_v = 0$

But, when y_{11} is missing and estimated by $\hat{x}$, although $E(MSE) = \sigma^2$,
$E(MS_{TR}(\hat{x})) = \sigma^2 + \{\text{constant}\} \times \sigma_t^2 + $ some positive expression
That is even under H_0,

$$E(MS_{TR}(\hat{x})) = \sigma^2 + \text{ some positive expression} > \sigma^2 = E(MSE)$$

Hence, $MS_{TR}(\hat{x})$ is not the proper valid error for testing H_0

Summary :

- If the approximate test **doesn't reject** H_0, there is no need to perform any accurate test of significance. Acceptance of the null hypothesis means $E(MS_{TR}(\hat{x}))$ is not significantly greater than the value of $E(MSE)$, so obviously $(E(MS_{TR}(\hat{x}))-$some positive number$)$ is not significantly greater than the value of $E(MSE)$ which means here the acceptance of the null hypothesis is certain.
- But, if the null hypothesis is **rejected**, we can't say whether the rejection is due to high value of the estimate of σ^2 or the other positive part. Thus we will conduct more accurate test.

2.1 More accurate testing procedure-1 when H_0 is rejected in the approximate test

- Here we use the estimated value, $\hat{x}$ only for calculation of SSE, same as in the Approximate testing procedure.
- But we don't use $\hat{x}$ in the calculation of SS_{BL} and TSS. SS_{TR} is calculated by subtraction.
- So in short, omitting $y_{11} = \hat{x}$ and assuming $\tau_1 = \tau_2 = ... = \tau_v$, we see that,
  - $SSE_{H_0} = SS_{TR} + SSE$ as given in this table is actually the SS due to error for such an experiment under H_0
  - Hence, $SS_{TR} = SS_{H_0} = SSE_{H_0} - SSE$ of this table is actually SS due to H_0 for this experiment.
  - The test statistic is $\frac{SS_{H_0}/(v-1)}{SSE(\hat{x})/(bv-v-b)} \overset{H_0}{\sim} F_{(v-1),(bv-b-v)}$

Sources	d.f	SS	MS
Blocks	$b-1$	$\frac{B_1'^2}{v-1} + \frac{\sum_{i=2}^{b} B_i^2}{v} - \frac{G'^2}{bv-1}$	MS_{BL}
Treatments	$v-1$	$\text{SS}_{TR} = TSS - SS_{BL} - SSE(\hat{x})$	MS_{TR}
Error	$(b-1)(v-1)-1$	As in the augmented ANOVA table $(SSE(\hat{x}))$	MSE
TSS	$bv-2$	$\sum_i \sum_j y_{ij}^2 - \frac{G'^2}{bv-1}$, $(i,j) \neq (1,1)$	—

Table 3: ANOVA table for More accurate testing procedure-1

2.2 More accurate testing procedure-2 when H_0 is rejected in the approximate test

Alternatively we may perform another more accurate test as follows,

We find least square estimate of y_{11}under $H_0(\tau_1 = \tau_2 = ... = \tau_v)$ say $\widetilde{x}$. This will be

$$\widetilde{x} = \frac{B_1'}{v-1}$$

Then SS due to H_0, $SS_{H_0} = SSE(\tilde{x}) - SSE(\hat{x})$, also called adjusted sum of squared due to treatment with $d.f. = v - 1$. The test statistic is given by,

$$\frac{SS_{H_0}/(v-1)}{SSE(\hat{x})/(bv-v-b)} \overset{H_0}{\sim} F_{(v-1),(bv-b-v)}$$

3 Accuracy of Approximations

Now we will see actually how much accurate these testing produces are.

METHODOLOGY : In RBD the model is,

$$y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \quad \text{where } \sum_i \beta_i = \sum_j \tau_j = 0 \;\&\; \epsilon_{ij} \overset{iid}{\sim} N(0,1)$$

- We will use simulation to generate sample observations from this model under H_0. But for that we have to fix some value of μ, β_i's and τ_j's .
- Under H_0 τ_j's are 0, so the model becomes : $y_{ij} = \mu + \beta_i + \epsilon$ where $\sum_i \beta_i = 0$ & $\epsilon_{ij} \overset{iid}{\sim} N(0,1)$
- Here we take number of blocks, $b = 5$ and number of treatments, $v = 6$.
- Also let, $\mu = 0$ and β_i's to be -2, -1, 0, 1, 2 . As $\sum_{i=1}^{5} \beta_i = 0$ we take such type of values of β_i's .
  Later we will investigate for more different type of values of μ and β_i's.
- Now our model is fixed and,

$$y_{ij} \sim N((i-3),1) \forall\, i = 1(1)5, j = 1(1)6 \quad ....(*)$$

- Then for different procedures we we will calculate the observed values of the test statistic, F_{obs} for each simulation and store it.
- Compare them with the approximate theoretical distribution and also within themselves.

A Measure of Accuracy of Approximation : The sum of absolute differences between empirical CDF and the theoretical CDF at the simulated points can be taken as a measure of accuracy of the actual distribution and the approximated F-distribution.

Note: Why this measure ?

Most common loss function is squared error loss function, which makes small values (modulus<1) more small and large values (modulus>1) more large. But observe that here these approximations are good enough and thus the errors are small which means if we use squared error loss, the order of accuracy may be not reflected through these measure. Thus to get an appropiate measure of comparison we take absolute error loss function as a measure of accuracy.

3.1 Algorithms :

- Fix the values of v, b, μ and β_i's . Here v and b are taken as 6 & 5 respectively.
- simulate values from $N(0, 1)$ and then take $y_{ij} = i - 3 +$ (simulated values) to get a set of simulated values of y_{ij}'s , where $i = 1(1)5$ and $j = 1(1)6$.

1. Approximate Test : by imputing the missing observation using its LSE
   - We calculate the values of B'_1,T'_1 & G'.
   - Estimate the missing value by $\hat{x}$.
   - Impute the estimated $\hat{x}$ in $(1, 1)^{\text{th}}$ cell.
   - Calculate $TSS, SS_{TR}, SS_{BL}, SSE$ from this imputed data.
   - Calculate observed value of test statistic $F_{obs} = \frac{MSTr}{MSE}$,
   - Repeat this many times and store it to a vector of observed F-values, "F_sim".
   - Plot histogram of simulated values of the test statistic.
   - Over the histogram plot the approximate distribution of the test statistic, which is here,$F_{6-1,(5-1)(6-1)-1} \equiv F_{5,19}$.
2. More accurate testing procedure-1 when $\boldsymbol{H_0}$ is rejected in the approximate test :
   - Use SSE as calculated above, $SSE(\hat{x}) = TSS(\hat{x}) - SS_{BL}(\hat{x}) - SS_{TR}(\hat{x})$.
   - Calculate SS_{BL} and TSS without using $\hat{x}$ as given in Table 2.
   - then By subtraction calculate SS_{TR}

$$SS_{TR} = TSS - SS_{BL} - SSE(\hat{x})$$

- Calculate observed value of test statistic $F_{obs} = \frac{MSTr}{MSE}$ and store it to a vector of observed F-values, "F_sim_1".
- Plot histogram of simulated values of the test statistic.
- Over the histogram plot the approximate distribution of the test statistic, which is here,$F_{6-1,(5-1)(6-1)-1} \equiv F_{5,19}$.

3. More accurate testing procedure-2 when $\boldsymbol{H_0}$ is rejected in the approximate test :

- Already we have $SSE(\hat{x}) = TSS(\hat{x}) - SS_{BL}(\hat{x}) - SS_{TR}(\hat{x})$ now,
  - Estimate the missing value by $\tilde{x}$ and impute it in $(1,1)^{\text{th}}$ cell instead of $\hat{x}$.
  - Calculate SSE using $\tilde{x}$, $SSE(\tilde{x}) = TSS(\tilde{x}) - SS_{BL}(\tilde{x}) - SS_{TR}(\tilde{x})$
- Calculate $SS_{H_0} = SSE(\tilde{x}) - SSE(\hat{x})$
- Calculate the value of the observed test statistic

$$F_{obs} = \frac{SS_{H_0}/(v-1)}{SSE(\hat{x})/(bv - v - b)}$$

  and store it to a vector of observed F-values, "F_sim_2"
- Plot histogram of simulated values of the test statistic
- Over the histogram plot the approximate distribution of the test statistic, which is here,$F_{6-1,(5-1)(6-1)-1} \equiv F_{5,19}$

R-code :

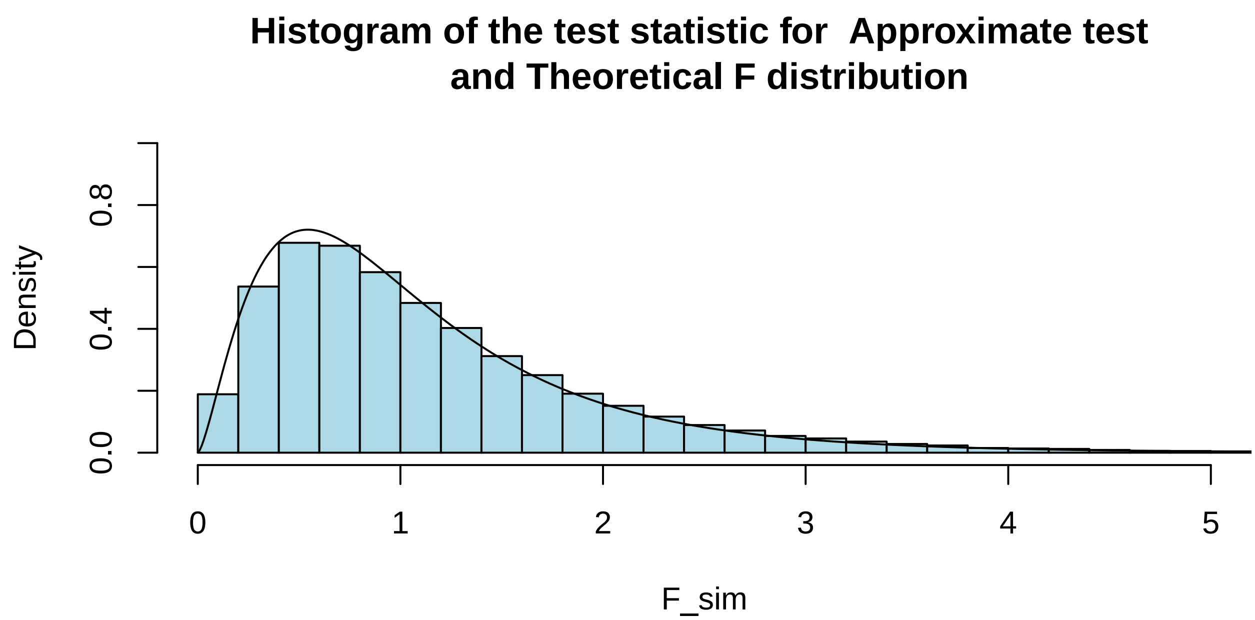
```
set.seed(1)
v = 6 ; b = 5
mu = 0 ; beta = seq(-(b-1)/2,(b-1)/2,length=5); tau = rep(0,v)
sim = 10^5

F_sim = NULL; F_sim_1 = NULL; F_sim_2 = NULL

for(i in 1:sim)
{
  data_sim = mu + outer(beta,tau,"+") + matrix(rnorm(b*v),nrow = b)
  data_sim[1,1] <- 0
```

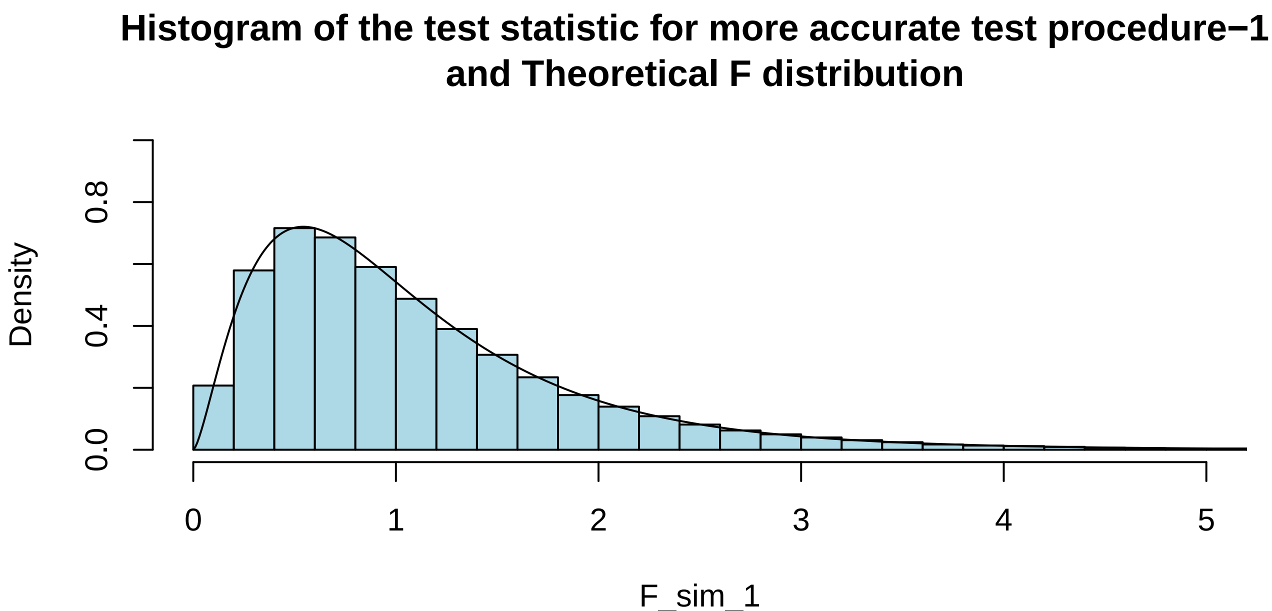
```
B.1_dash <- sum(data_sim[1,])          # B.1_dash is vector of B1'
T.1_dash <- sum(data_sim[,1])          # T.1_dash is vector of T1'
G_dash <- sum(data_sim)                # G_dash is vector of G'

x_hat <- (b*B.1_dash+v*T.1_dash-G_dash)/((b-1)*(v-1))
x_curl <- B.1_dash/(v-1)

                 #### approximate test ####

data_sim[1,1] <- x_hat
B <- rowSums(data_sim)
T <- colSums(data_sim)
G <- sum(data_sim)
SSBl = sum(B^2)/v - (G^2/(b*v))
SSTr = sum(T^2)/b - (G^2/(b*v)) ; MSTr = SSTr/(v-1)
TSS = sum(data_sim^2) - (G^2/(b*v))
SSE = TSS - SSBl - SSTr ; MSE = SSE/(b*v-b-v)
F_sim[i] <- MSTr/MSE

sse_hat=SSE

          #### More accurate testing procedure-1  ####

SSBl = (B.1_dash^2)/(v-1) + (sum(B[-1]^2)/v) - (G_dash^2/(b*v - 1))
TSS = sum(data_sim^2)-data_sim[1,1]^2 - (G_dash^2/(b*v - 1))
SSTr = TSS - SSBl - sse_hat ;  MSTr = SSTr/(v-1)
F_sim_1[i] <- MSTr/(sse_hat/(b*v-b-v))

          #### More accurate testing procedure-2  ####

data_sim[1,1] <- x_curl
B <- rowSums(data_sim)
T <- colSums(data_sim)
G <- sum(data_sim)

SSBl = sum(B^2)/v - (G^2/(b*v))
SSTr = sum(T^2)/b - (G^2/(b*v))
```

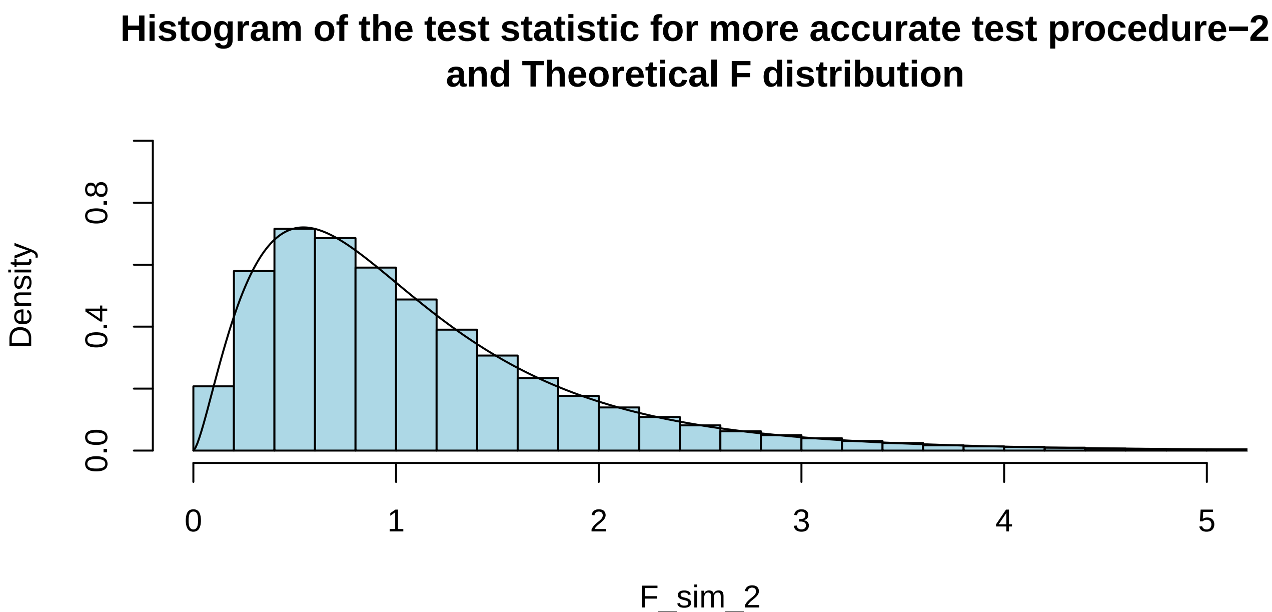
```
  TSS = sum(data_sim^2) - (G^2/(b*v))
  SSE0 = TSS - SSBl
  F_sim_2[i] <- ((SSE0 - SSE)/(v-1))/(sse_hat/(b*v-b-v))
}
```

Plots :

Now, we shall plot the empirical histograms along with the theoretical curve of $F_{v-1,bv-b-v}$ to study which of the test statistics is/are closest to $F_{v-1,bv-b-v}$ under H_0.



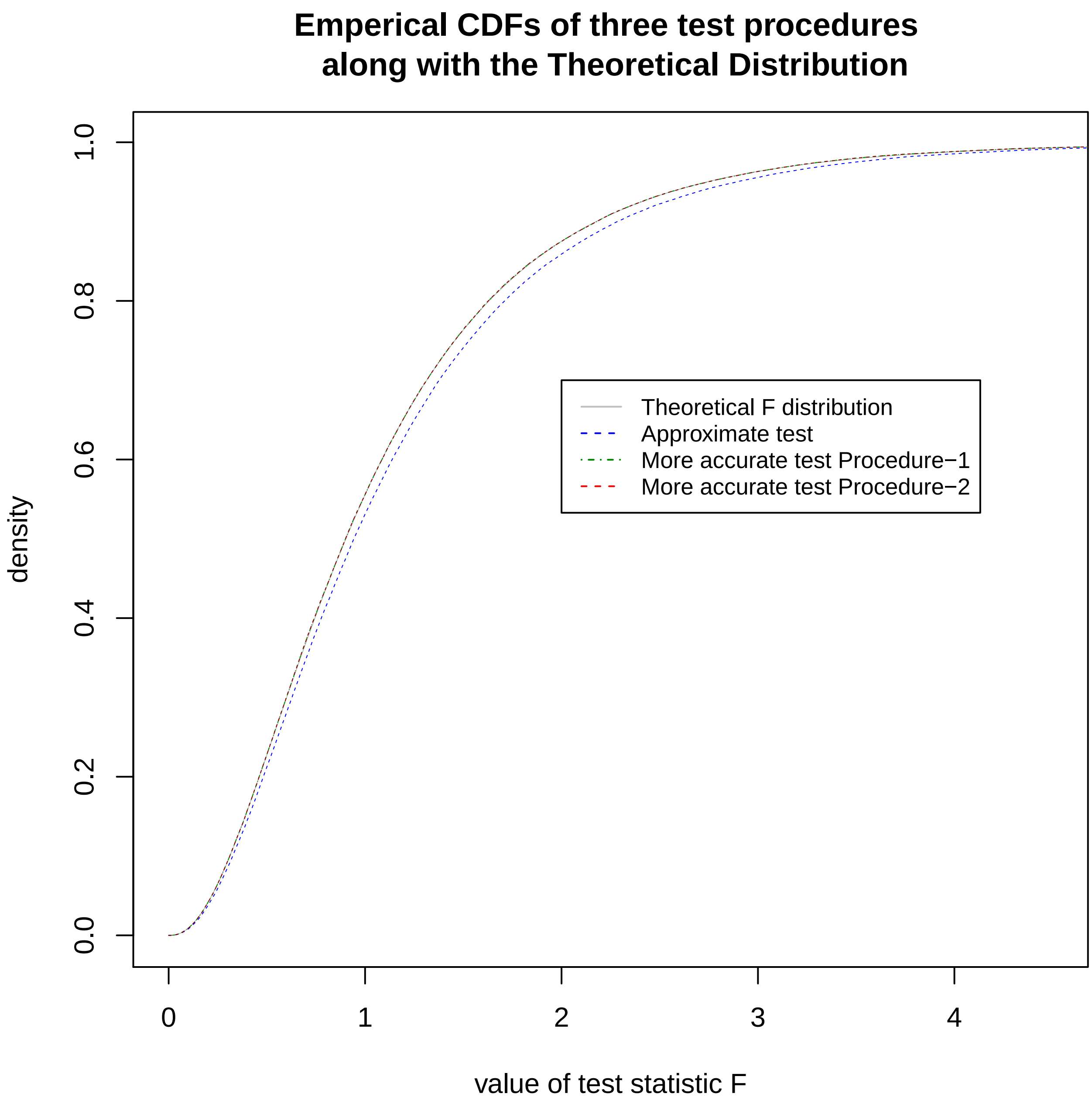




- So these distributions are actually very close to the theoretical F-distribution

To compare them together, we will now plot all the empirical CDFs together along with the theoretical CDF i.e. of $F_{v-1,bv-b-v}$.

```
sup = seq(0,6,0.01)
F_ecdf = ecdf(F_sim)
F_ecdf_1 = ecdf(F_sim_1)
F_ecdf_2 = ecdf(F_sim_2)

plot(sup,pf(sup,df1=v-1,df2=b*v-b-v),type = "l",lwd=0.5,col="grey",xlim=c(0,4.5),
  xlab="value of test statistic F", ylab="density",
main="Emperical CDFs of three test procedures \n along with the Theoretical Distribution")
lines(sup,F_ecdf(sup),type = "l",lwd=0.5,lty=2,col="blue")
lines(sup,F_ecdf_1(sup),type = "l",lwd=0.5,lty=4,col="green4")
lines(sup,F_ecdf_2(sup),type = "l",lwd=0.5,lty=2,col="red4")
legend(2,0.7, legend=c("Theoretical F distribution","Approximate test",
 "More accurate test Procedure-1","More accurate test Procedure-2"),
  col=c("grey","blue","green4","red"), lty=c(1,2,4,2), cex=0.8)
```



We can see that the distribution of the simulated Test statistics are a little-bit deviated from the approximate theoretical distribution in Approximate test. But more accurate test procedures 1 & 2 are almost similar to the approximate theoretical distribution, $F_{v-1,bv-b-v}$.

- As defined earlier, here the measures of accuracy of approximation of all three procedures i.e. the absolute difference from theoretical CDF those are respectively given by,

```
aprox_F=pf(sup,df1 = (v-1),df2 = b*v-b-v)
sum(abs(F_ecdf(sup)-aprox_F))

[1] 5.350146

sum(abs(F_ecdf_1(sup)-aprox_F))

[1] 0.2229855

sum(abs(F_ecdf_2(sup)-aprox_F))

[1] 0.2229855
```

- See that that the measure of accuracy is exactly equal for last two procedures. Now there can be doubt about this procedure. **How can they be exactly equal ?**
  - Actually these two vectors "F_sim_1" and "F_sim_2" are not exactly same but they are too close. To establish the fact observe that,

  ```
  sum(abs(F_sim_1-F_sim_2))

  [1] 1.851524e-10
  ```

  That is the sum of the absolute differences are not exactly 0 but the difference is too small i.e. 1.851524×10^{-10}.

4 Size

The empirical sizes of the tests are proportion of times the test rejects the null hypothesis when it is actually true. So we will consider the proportion of times the simulated value of the the F statistic is bigger than the critical point.

```
alpha=0.05
cut_pt = qf(p = alpha ,df1 = (v-1),df2 = b*v-b-v,lower.tail = FALSE)
mean(F_sim > cut_pt)

[1] 0.05903

mean(F_sim_1 > cut_pt)

[1] 0.05037

mean(F_sim_2 > cut_pt)

[1] 0.05037
```

Interpretation : The Approximate test which gives $size = 0.05903 > 0.05$ significantly i.e. it rejects more often, even when the null hypothesis is true. But the other two testing procedures are more accurate as we see their empirical sizes are very close to 0.05.

5 Power Curves of Different Testing procedures

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_v = 0$$
$$vs$$
$$H_1 : \text{ at least one of the values are different from 0}$$

$$\begin{aligned}\text{So power } &= P_{H_1}\left(F_{obs} > F_{0.05;(v-1),(bv-b-v)}\right) \\ &= [\beta(\tau)]_{\{\text{at least one of the} \tau_i \text{ is different from 0 }\}}\end{aligned}$$

Here observe that , as $\sum \tau_i^2$ increases we will tend to reject the Null hypothesis. Thus we will find the power for different values of $\sum \tau_i^2$.

5.1 Algorithm :

- Fix the values of v, b, μ and β_i's . Here v and b are taken as 6 & 5 respectively and β_i's are chosen randomly from, $U(-b, b)$ and then adjusted by, $\beta_i - \bar{\beta}$, so that $\sum \beta_i = 0$.

- Define a function which calculates the estimates and the sum of squares in the similar manner as discussed in **section 3.1**

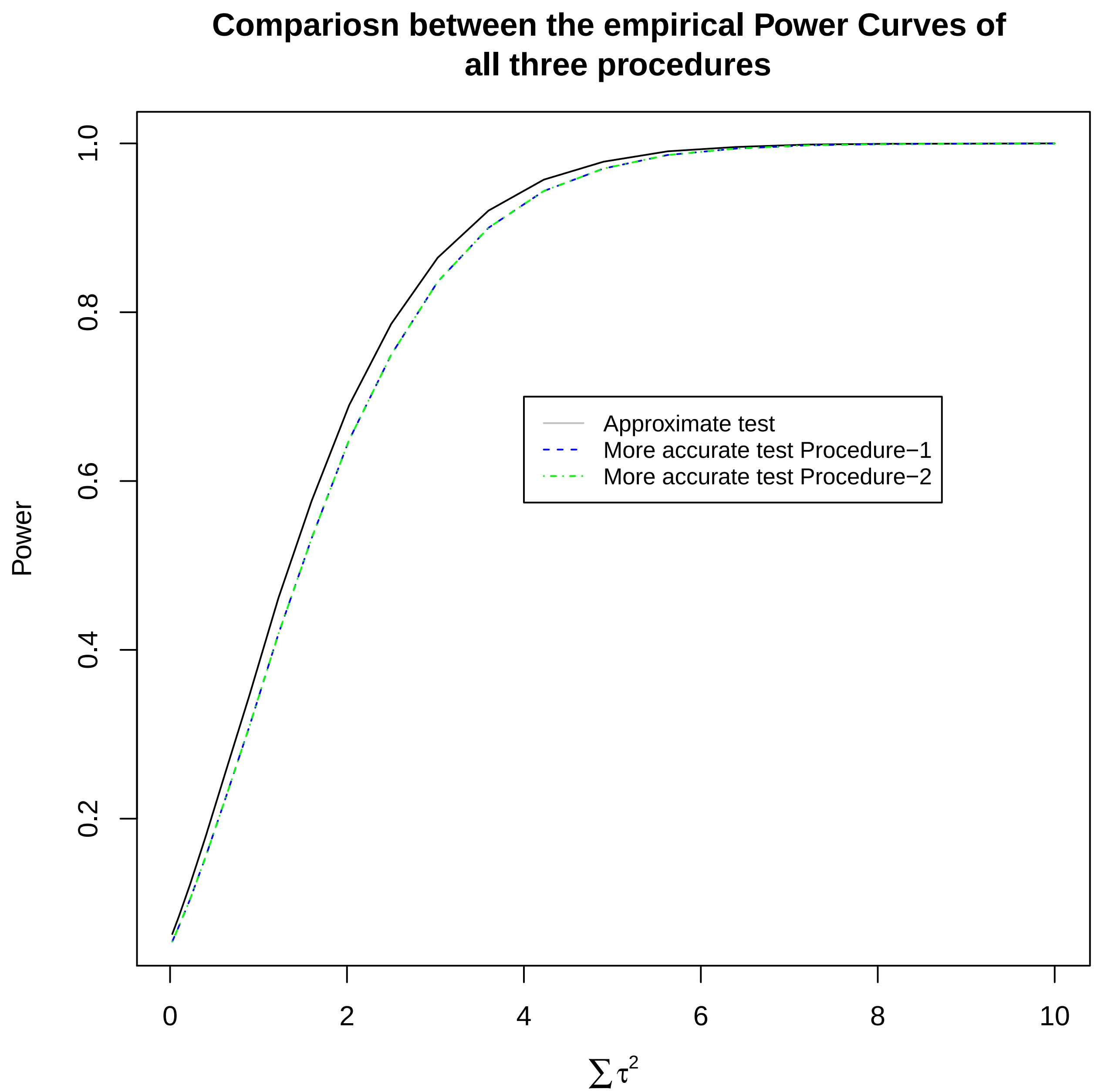
  1. For each sample note that if, $F_{obs} >$ the cut point and store that TRUE/FALSE outcome stored in vectors : "power.vec" , "power.vec_1" & "power.vec_2".
  2. Then the mean of these vectors are actually the proportion of times the observed F statistic falls into the critical region which is the empirically calculated power of the test for a specified value of τ_i's as alternative hypothesis.

R-code :

```
set.seed(50)
sim = 10^5; alpha = 0.05
v0 = 5; b0 = 6
beta0 = runif(b0,-b0,b0)
beta0 = beta0 - mean(beta0)
Power_Curve=function(tau=tau0,v=v0,b=b0,mu=0,beta=beta0)
{
  power.vec = NULL; power.vec_1 = NULL; power.vec_2 = NULL
  cut_pt = qf(p = alpha,df1 = (v-1),df2 = (b-1)*(v-1) - 1,lower.tail = FALSE)
  for(i in 1:sim)
  {
   data_sim = mu + outer(beta,tau,"+") + matrix(rnorm(b*v),nrow = b)
   data_sim[1,1] <- 0

   B.1_dash <- sum(data_sim[1,])          # B.1_dash is vector of B1'
   T.1_dash <- sum(data_sim[,1])          # T.1_dash is vector of T1'
   G_dash <- sum(data_sim)                # G_dash is vector of G'

   x_hat <- (b*B.1_dash+v*T.1_dash-G_dash)/((b-1)*(v-1))
   x_curl <- B.1_dash/(v-1)
```

```
                  #### approximate test ####

   data_sim[1,1] <- x_hat
   B <- rowSums(data_sim)
   T <- colSums(data_sim)
   G <- sum(data_sim)
   SSBl = sum(B^2)/v - (G^2/(b*v))
   SSTr = sum(T^2)/b - (G^2/(b*v)) ; MSTr = SSTr/(v-1)
   TSS = sum(data_sim^2) - (G^2/(b*v))
   SSE = TSS - SSBl - SSTr ; MSE = SSE/(b*v-b-v)
   power.vec[i] = (MSTr/MSE > cut_pt)

   sse_hat=SSE

           #### More accurate testing procedure-1  ####

   SSBl = (B.1_dash^2)/(v-1) + (sum(B[-1]^2)/v) - (G_dash^2/(b*v - 1))
   TSS = sum(data_sim^2)-data_sim[1,1]^2 - (G_dash^2/(b*v - 1))
   SSTr = TSS - SSBl - sse_hat ;  MSTr = SSTr/(v-1)
   power.vec_1[i] = (MSTr/(sse_hat/(b*v-b-v)) > cut_pt)

           #### More accurate testing procedure-2  ####

   data_sim[1,1] <- x_curl
   B <- rowSums(data_sim)
   T <- colSums(data_sim)
   G <- sum(data_sim)

   SSBl = sum(B^2)/v - (G^2/(b*v))
   SSTr = sum(T^2)/b - (G^2/(b*v))
   TSS = sum(data_sim^2) - (G^2/(b*v))
   SSE0 = TSS - SSBl
   power.vec_2[i] = (((SSE0 - SSE)/(v-1))/(sse_hat/(b*v-b-v)) > cut_pt)
  }
  power=matrix(c(power.vec,power.vec_1,power.vec_2),nrow=sim)
}
```

```r
n = 20
sample.tau = NULL
tau_mat = NULL

for(i in 1:n)
{
  sample.tau = seq(-i/10,i/10,length = v0)
  tau_mat = rbind(tau_mat,sample.tau -  mean(sample.tau))
}

sum_tau_sq = apply(tau_mat,MARGIN = 1,FUN = function(x){return(sum(x^2))})

Power = NULL; Power_1 = NULL; Power_2 = NULL
for(i in 1:n)
{
  mat=Power_Curve(tau = tau_mat[i,])
  Power[i] = colMeans(mat)[1]
  Power_1[i] = colMeans(mat)[2]
  Power_2[i] = colMeans(mat)[3]
}

plot(sum_tau_sq,Power,xlab = expression(sum(tau^2)),ylab="Power",type = "l",
    main="Compariosn between the empirical Power Curves of \n all three procedures")
lines(sum_tau_sq,Power_1,type = "l",lty=2,col="blue")
lines(sum_tau_sq,Power_2,type = "l",lty=4,col="green")
legend(4,0.7, legend=c("Approximate test","More accurate test Procedure-1",
       "More accurate test Procedure-2"),col=c("grey","blue","green"),
       lty=c(1,2,4), cex=0.8)
```



- Here also we can see that the power curve of approximate test lies over the curves of more accurate tests, which indicates, it rejects the null hypothesis more often than the accurate tests when the null hypothesis is actually false.

6 Comparisons for Different Values of μ, β_i's and σ

As the missing data is estimated and that estimate is such that the usual orthogonality of treatment effects and block effects is lost, we may think that the values of μ and β_i's may affect our previous findings. Also we can check the consistency of our findings

Procedure :

1. we will consider different examples where we choose
   - $\mu = 2$ or, 5.
   - β_i's are randomly chosen such that, $\sum \beta_i = 0$
   - we take two different values of $\sigma = 2$ or 4

   we also change the **seed** in the code
2. In each cases we will calculate the measures of accuracy and the size of the test.

μ	σ	β_i's	measure of accuracy for 3 tests			Size for 3 tests		
			1	2	3	1	2	3
2	2	-1.66, 4.86, 4.50, 1.21,-8.91	6.063186	0.5259657	0.5259657	0.06017	0.05134	0.05134
		-4.16, -0.41, 2.56, 2.15, -0.14	5.135479	0.3649022	0.3649022	0.058	0.04981	0.04981
	5	2.96, -2.43, -3.59, -2.8, 5.86	5.373459	0.303278	0.303278	0.0598	0.05082	0.05082
		2.14, -2.81, 2.73, 1.60, -3.66	5.854466	0.3583031	0.3583031	0.05946	0.05045	0.05045
5	2	2.35, -2.69, 4.83, 4.69, -9.18	5.253546	0.2912271	0.2912271	0.05841	0.04963	0.04963
		-4.09, 3.68, 4.81, -0.95, -3.45	5.133166	0.3622729	0.3622729	0.05795	0.0489	0.0489
	5	1.63, 4.25, -4.12, 4.04, -5.8	5.677847	0.2955843	0.2955843	0.05944	0.0502	0.0502
		0.81, -0.58, 3.34, 3.17, -6.74	5.305436	0.2646031	0.2646031	0.05818	0.04935	0.04935
4	15	18, -15, 13,21, -37	5.684606	0.4339811	0.4339811	0.05835	0.04954	0.04954
	17	25,11,-16,-6,-14	5.614976	0.2090806	0.2090806	0.05955	0.05028	0.05028

Table 4: Measure of accuracy and size for different values of μ , β_i's and σ

7 Conclusion

- We conclude that these approximate test procedures are satisfactory and truly reflects the actual scenario.
- These two more accurate testing procedures performs similar and better than the approximate test on an average.
- Hence for any such one missing plot data in a RBD
  - we will conduct the approximate test first
    - if it <u>accepts</u> the Null Hypothesis then we can say treatment effects are all equal i.e. $\tau_j = 0, \forall j$ and there is <u>no need to perform more accurate tests</u>.
    - if it <u>rejects</u> the Null Hypothesis then we will conduct any of the <u>more accurate tests for ultimate conclusion.</u>

Bibliography :

[1] Mukhopadhyay Parimal (2005); Applied Statistics (2nd. Ed); Kolkata Books and Allied Pvt. Ltd.

[2] Dasgupta B., Gun A. M., Gupta M. K. (2008); Fundamentals of Statistics Vol-2 (9th. Ed); World Press

† * * * * THE END * * * * †