

# Multiple Comparison :R Lab (Fisher's LSD)

Nithya Dev

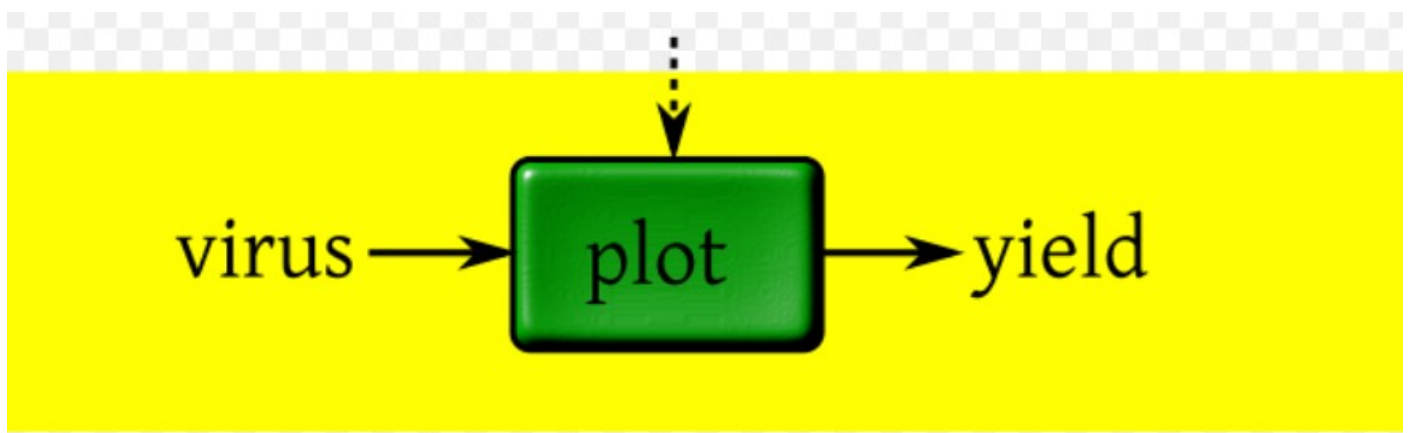
BS2018

## Multiple Comparison Using R:

It is not a pleasant experience to do multiple comparison testing in R because base R doesn't support much of multiple comparison. We will need additional packages specifically *agricolae*

Let's see the type of data we are going to work with and how to perform Fisher's LSD

Consider an example where we have plots of sweet potato and there are a couple of viruses that are being fed into it.



We have two viruses so four possible combinations which are Both Virus 1 and Virus 2 present , Both Virus 1 and Virus 2 absent, Virus 1 present and Virus 2 absent , Virus 1

absent and Virus 2 present. So we virus has got 4 levels, it's a factor input and are measuring the yield of sweet potatoes. It's an one way ANOVA setup.

Let's see how to do this in R

```
> install.packages("agricolae") #Installing the agricolae package
> library(agricolae) #loading the package
> data(sweetpotato) #This dataset is part of the package and for some reason if we don't load this then we can't work with that dataset
> summary(sweetpotato)
```

Output:

```
virus      yield
cc:3      Min.      :10.60
fc:3      1st Qu.    :20.00
ff:3      Median    :28.25
oo:3      Mean      :27.62
           3rd Qu.   :38.45
           Max.     :41.80
```

cc:3 → only virus c is present

fc:3 → virus c and virus f is present

ff:3 → only virus f is present

oo:3 → both viruses are absent

Everything is repeated thrice.

Now we will perform Fisher's LSD which has two steps. In first step we will carry out the usual ANOVA test.

```
> model=lm(yield~virus,data=sweetpotato) #fitting the data
> model0=lm(yield~1,data=sweetpotato)    #fitting the null model
> anova(model0,model)
```

Output:

Analysis of Variance Table

Model 1: yield ~ 1

Model 2: yield ~ virus

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	11	1350.12				
2	8	179.91	3	1170.2	17.345	0.0007334 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

We find that the p-value is very small and highly significant(\*\*\*).So we have passed first test of Fisher's LSD.

Now,

```
> out=LSD.test(model,"virus",group=F) #LSD.test is included in the package
> out
```

We get a lot of things in the output in which we focus on the comparison part.

Output:

```
$comparison
      difference pvalue signif.      LCL      UCL
cc - fc  11.5333333 0.0176      *    2.604368  20.462299
cc - ff -11.9333333 0.0151      *   -20.862299  -3.004368
cc - oo -12.5000000 0.0121      *   -21.428965  -3.571035
fc - ff -23.4666667 0.0003     ***   -32.395632 -14.537701
fc - oo -24.0333333 0.0003     ***   -32.962299 -15.104368
ff - oo  -0.5666667 0.8873           -9.495632   8.362299
```

If we are working at 5% significance level then we accept  $ff-oo$  and everything else is significant, which is one way of drawing conclusion.

Also by using the above command we can make grouping if we take  $group=T$  in the command ,but this new command won't provide us the comparison we got earlier

```
> (out=LSD.test(model,"virus",group=T))
```

Output:

```
$comparison
NULL

$groups
      yield groups
oo 36.90000      a
ff 36.33333      a
cc 24.40000      b
fc 12.86667      c
```

we can see that both oo and ff are in same group but cc and fc are in different groups.

Now we have a problem here because exact equality is transitive but approximate equality is not. So it may happen that we say level 1 and level 2 are approximately same, level 2 and level 3 are approximately same but that doesn't mean that level 3 and level 1 are approximately same.

To tackle this we take the Bonferroni test.