Linear Models Assignment 3

Subhrajyoty Roy (BS-1613)

October 25, 2018

Loading the Prequisites Libraries

The package *reshape2* has been used in order to mainpulate the structure of the data and transform it into the required format. I have mainly used *DescTools* package to obtain the multiple comparison tests. However, there is no support for bonferroni test for arbitrary contrasts in *DescTools* package, so I need to use *multcomp* package for this purpose. The guide is given in the url: https://stat.ethz.ch/~meier/teaching/anova/contrasts-and-multiple-testing.html

```
library(reshape2)
library(DescTools)
library(multcomp)
```

I have used the packages *rmarkdown*, *knitr* and *kableExtra* in order to generate the report in .pdf format.

```
library(knitr)
library(kableExtra)
```

Question 1

Consider problem 1 of your Assignment 1. For each program, you delete two observations and thus it is a balanced one way model.

```
firmdata <- data.frame(prog1 = c(9,12,14,11,13), prog2 = c(10,6,9,9,10),
                       prog3 = c(12,14,11,13,11), prog4 = c(9,8,11,7,8))
kable(firmdata) %>% kable_styling(full_width = F)
```

prog1	prog2	prog3	prog4
9	10	12	9
12	6	14	8
14	9	11	11
11	9	13	7
13	10	11	8

Now, I have to delete two observations for each program.

```
set.seed(1613)  #this is my roll number, set the seed for reproducbility
x = sample(1:5, size = 2)
firmdata = firmdata[-x,]  #remove the corresponding observations
firmdata = melt(firmdata)
firmdata$variable = as.factor(firmdata$variable)
kable(firmdata) %>% kable_styling(full_width = F)
```

variable	value
prog1	9
prog1	12
prog1	13
prog2	10
prog2	6
prog2	10
prog3	12
prog3	14
prog3	11
prog4	9
prog4	8
prog4	8

Let the effect of Program i be denoted by α_i. Let the set S1 denote the collection of elementary contrasts of the form $\alpha_i - \alpha_j, i < j$. Let S2 denote the collection of general contrasts of the form $\alpha_1 - 2\alpha_2 + \alpha_3$ and $2\alpha_1 - \alpha_2 - \alpha_3$. Let S3 be the union of S2 and S1. Compute simultaneous confidence intervals with 95% confidence coefficient for S1 with Scheffe, Tukey and Bonferroni's method.

```
#fits the ANOVA model
model = aov(value ~ variable, firmdata)

#calculates the Tukey HSD
res = PostHocTest(model, method = "hsd", conf.level = 0.95)
res[[1]] = cbind(res[[1]], length = (res[[1]][,"upr.ci"] - res[[1]][,"lwr.ci"]))
#create a new column containing length of the intervals
print(res)
```

```
  Posthoc multiple comparisons of means : Tukey HSD
    95% family-wise confidence level

$variable
                  diff     lwr.ci    upr.ci   pval   length
prog2-prog1 -2.6666667 -7.2579456 1.9246123 0.3155 9.182558
prog3-prog1  1.0000000 -3.5912789 5.5912789 0.8952 9.182558
prog4-prog1 -3.0000000 -7.5912789 1.5912789 0.2340 9.182558
prog3-prog2  3.6666667 -0.9246123 8.2579456 0.1241 9.182558
prog4-prog2 -0.3333333 -4.9246123 4.2579456 0.9952 9.182558
prog4-prog3 -4.0000000 -8.5912789 0.5912789 0.0895 9.182558 .

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

```
#calculates the Scheffe's test
res = PostHocTest(model, method = "scheffe", conf.level = 0.95)
res[[1]] = cbind(res[[1]], length = (res[[1]][,"upr.ci"] - res[[1]][,"lwr.ci"]))
print(res)
```

```
  Posthoc multiple comparisons of means : Scheffe Test
    95% family-wise confidence level

$variable
```

```
                 diff    lwr.ci   upr.ci   pval   length
prog2-prog1 -2.6666667 -7.674139 2.340806 0.3854 10.01494
prog3-prog1  1.0000000 -4.007472 6.007472 0.9188 10.01494
prog4-prog1 -3.0000000 -8.007472 2.007472 0.2967 10.01494
prog3-prog2  3.6666667 -1.340806 8.674139 0.1682 10.01494
prog4-prog2 -0.3333333 -5.340806 4.674139 0.9964 10.01494
prog4-prog3 -4.0000000 -9.007472 1.007472 0.1249 10.01494

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

```r
#calculates the Bonferroni's test
res = PostHocTest(model, method = "bonferroni", conf.level = 0.95)
res[[1]] = cbind(res[[1]], length = (res[[1]][,"upr.ci"] - res[[1]][,"lwr.ci"]))
print(res)
```

```
  Posthoc multiple comparisons of means : Bonferroni
    95% family-wise confidence level

$variable
                 diff    lwr.ci    upr.ci   pval   length
prog2-prog1 -2.6666667 -7.654408 2.3210751 0.5996 9.975483
prog3-prog1  1.0000000 -3.987742 5.9877417 1.0000 9.975483
prog4-prog1 -3.0000000 -7.987742 1.9877417 0.4185 9.975483
prog3-prog2  3.6666667 -1.321075 8.6544084 0.2027 9.975483
prog4-prog2 -0.3333333 -5.321075 4.6544084 1.0000 9.975483
prog4-prog3 -4.0000000 -8.987742 0.9877417 0.1414 9.975483

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

For S2 and S3 obtain simultaneous confidence intervals using Scheffe's method as well as Bonferroni's method.

Firstly, the matrix containing the coefficients for each of the contrasts is created.

```r
s2 = Permn(c(0,-1,-1,2))  #computes all permutations of the coefficients
colnames(s2) <-paste("Coefficient of prog", 1:4)
kable(s2) %>% kable_styling(full_width = F)
```

Coefficient of prog 1	Coefficient of prog 2	Coefficient of prog 3	Coefficient of prog 4
0	-1	-1	2
-1	0	-1	2
-1	-1	0	2
-1	-1	2	0
0	-1	2	-1
-1	0	2	-1
-1	2	0	-1
-1	2	-1	0
0	2	-1	-1
2	0	-1	-1
2	-1	0	-1
2	-1	-1	0

I obtain simulteneous confidence intervals for the above set of contrasts in S2 using the method of Scheffe.

```
res = ScheffeTest(model, conf.level = 0.95, contrasts = t(s2))
res[[1]] = cbind(res[[1]], length = (res[[1]][,"upr.ci"] - res[[1]][,"lwr.ci"]))
print(res)
```

```

  Posthoc multiple comparisons of means : Scheffe Test
    95% family-wise confidence level

$variable
                       diff     lwr.ci    upr.ci   pval   length
prog4-prog2,prog3 -4.333333 -13.006530  4.339863 0.4353 17.34639
prog4-prog1,prog3 -7.000000 -15.673197  1.673197 0.1204 17.34639
prog4-prog1,prog2 -3.333333 -12.006530  5.339863 0.6325 17.34639
prog3-prog1,prog2  4.666667  -4.006530 13.339863 0.3774 17.34639
prog3-prog2,prog4  7.666667  -1.006530 16.339863 0.0849 17.34639 .
prog3-prog1,prog4  5.000000  -3.673197 13.673197 0.3250 17.34639
prog2-prog1,prog4 -2.333333 -11.006530  6.339863 0.8286 17.34639
prog2-prog1,prog3 -6.333333 -15.006530  2.339863 0.1697 17.34639
prog2-prog3,prog4 -3.333333 -12.006530  5.339863 0.6325 17.34639
prog1-prog3,prog4  2.000000  -6.673197 10.673197 0.8824 17.34639
prog1-prog2,prog4  5.666667  -3.006530 14.339863 0.2368 17.34639
prog1-prog2,prog3  1.666667  -7.006530 10.339863 0.9267 17.34639

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

I obtain simulteneous confidence intervals for the same set of contrasts using the method of Bonferroni.

```
model.gh = glht(model, linfct = mcp(variable = s2))  #set the contrasts
res.gh = confint(model.gh, test = adjusted("bonferroni"))
print(res.gh)
```

```

     Simultaneous Confidence Intervals

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = value ~ variable, data = firmdata)

Quantile = 3.3277
95% family-wise confidence level


Linear Hypotheses:
        Estimate lwr      upr
1 == 0   -4.3333 -12.5968   3.9302
2 == 0   -7.0000 -15.2635   1.2635
3 == 0   -3.3333 -11.5968   4.9302
4 == 0    4.6667  -3.5968  12.9302
5 == 0    7.6667  -0.5968  15.9302
6 == 0    5.0000  -3.2635  13.2635
7 == 0   -2.3333 -10.5968   5.9302
8 == 0   -6.3333 -14.5968   1.9302
```

```
9 == 0   -3.3333 -11.5968   4.9302
10 == 0   2.0000  -6.2635  10.2635
11 == 0   5.6667  -2.5968  13.9302
12 == 0   1.6667  -6.5968   9.9302
```

```r
res.gh$confint[,"upr"] - res.gh$confint[,"lwr"] #compute the lengths of CI's
```

```
     1      2      3      4      5      6      7      8      9     10
16.527 16.527 16.527 16.527 16.527 16.527 16.527 16.527 16.527 16.527
    11     12
16.527 16.527
```

Now, I have performed the similar for the set of contrasts in s3. There would be (12+6) = 18 many contrasts in s3.

```r
#computes the 6 pairwise contrasts
s1 <- rbind(c(-1,1,0,0), c(-1,0,1,0), c(-1,0,0,1), c(0,-1,1,0), c(0,-1,0,1), c(0,0,-1,1))
s3 <- rbind(s1, s2)
colnames(s3) <-paste("Coefficient of prog", 1:4)
kable(s3) %>% kable_styling(full_width = F)
```

Coefficient of prog 1	Coefficient of prog 2	Coefficient of prog 3	Coefficient of prog 4
-1	1	0	0
-1	0	1	0
-1	0	0	1
0	-1	1	0
0	-1	0	1
0	0	-1	1
0	-1	-1	2
-1	0	-1	2
-1	-1	0	2
-1	-1	2	0
0	-1	2	-1
-1	0	2	-1
-1	2	0	-1
-1	2	-1	0
0	2	-1	-1
2	0	-1	-1
2	-1	0	-1
2	-1	-1	0

```r
res = ScheffeTest(model, conf.level = 0.95, contrasts = t(s3))
res[[1]] = cbind(res[[1]], length = (res[[1]][,"upr.ci"] - res[[1]][,"lwr.ci"]))
print(res)
```

```

  Posthoc multiple comparisons of means : Scheffe Test
    95% family-wise confidence level

$variable
                        diff     lwr.ci    upr.ci   pval   length
prog2-prog1       -2.6666667  -7.674139  2.340806 0.3854 10.01494
prog3-prog1        1.0000000  -4.007472  6.007472 0.9188 10.01494
prog4-prog1       -3.0000000  -8.007472  2.007472 0.2967 10.01494
prog3-prog2        3.6666667  -1.340806  8.674139 0.1682 10.01494
```

```
prog4-prog2       -0.3333333  -5.340806  4.674139 0.9964 10.01494
prog4-prog3       -4.0000000  -9.007472  1.007472 0.1249 10.01494
prog4-prog2,prog3 -4.3333333 -13.006530  4.339863 0.4353 17.34639
prog4-prog1,prog3 -7.0000000 -15.673197  1.673197 0.1204 17.34639
prog4-prog1,prog2 -3.3333333 -12.006530  5.339863 0.6325 17.34639
prog3-prog1,prog2  4.6666667  -4.006530 13.339863 0.3774 17.34639
prog3-prog2,prog4  7.6666667  -1.006530 16.339863 0.0849 17.34639 .
prog3-prog1,prog4  5.0000000  -3.673197 13.673197 0.3250 17.34639
prog2-prog1,prog4 -2.3333333 -11.006530  6.339863 0.8286 17.34639
prog2-prog1,prog3 -6.3333333 -15.006530  2.339863 0.1697 17.34639
prog2-prog3,prog4 -3.3333333 -12.006530  5.339863 0.6325 17.34639
prog1-prog3,prog4  2.0000000  -6.673197 10.673197 0.8824 17.34639
prog1-prog2,prog4  5.6666667  -3.006530 14.339863 0.2368 17.34639
prog1-prog2,prog3  1.6666667  -7.006530 10.339863 0.9267 17.34639

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

```r
model.gh = glht(model, linfct = mcp(variable = s3))  #set the contrasts
res.gh = confint(model.gh, test = adjusted("bonferroni"))
print(res.gh)
```

```
     Simultaneous Confidence Intervals

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = value ~ variable, data = firmdata)

Quantile = 3.369
95% family-wise confidence level


Linear Hypotheses:
        Estimate lwr      upr
1 == 0   -2.6667  -7.4968   2.1635
2 == 0    1.0000  -3.8301   5.8301
3 == 0   -3.0000  -7.8301   1.8301
4 == 0    3.6667  -1.1635   8.4968
5 == 0   -0.3333  -5.1635   4.4968
6 == 0   -4.0000  -8.8301   0.8301
7 == 0   -4.3333 -12.6994   4.0327
8 == 0   -7.0000 -15.3661   1.3661
9 == 0   -3.3333 -11.6994   5.0327
10 == 0   4.6667  -3.6994  13.0327
11 == 0   7.6667  -0.6994  16.0327
12 == 0   5.0000  -3.3661  13.3661
13 == 0  -2.3333 -10.6994   6.0327
14 == 0  -6.3333 -14.6994   2.0327
15 == 0  -3.3333 -11.6994   5.0327
16 == 0   2.0000  -6.3661  10.3661
17 == 0   5.6667  -2.6994  14.0327
18 == 0   1.6667  -6.6994  10.0327
```

```r
res.gh$confint[,"upr"] - res.gh$confint[,"lwr"] #compute the lengths of CI's
```

```
        1         2         3         4         5         6         7 
 9.660288  9.660288  9.660288  9.660288  9.660288  9.660288 16.732110 
        8         9        10        11        12        13        14 
16.732110 16.732110 16.732110 16.732110 16.732110 16.732110 16.732110 
       15        16        17        18 
16.732110 16.732110 16.732110 16.732110 
```

For each of the three sets, which procedure works out the best?

The discussion regarding the best procedure has been divided into the three parts, each for different sets of contrasts.

1. Starting with S1, Tukey's Honest Significant Difference method gives a 95% confidence interval of length 9.18 units, Scheffe's method gives a 95% CI of length 10.01 units and Bonferroni's method gives a 95% CI of length 9.97 units. Clearly, from the tightness of the confidence interval as the evaluation criterion, it is seen that Tukey's method works best for pairwise contrasts.
2. For the set of contrasts in S2, Scheffe's method gives the 95% confidence interval of length 17.34 units, while Bonferroni's method gives the 95% CI of length 16.52 units. In this regard, Bonferroni's method works better.
3. For the set of contrasts in S2, Scheffe's method gives the 95% confidence interval of length 10.01 units for the pairwise difference contrasts and 17.34 units for the general contrasts, while Bonferroni's method gives the 95% CI of length 9.66 units for pairwise difference contrasts and 16.73 units for the general contrasts. In this case also, Bonferroni's method works better.

Finally, we see that Tukey's test is best in case of finding the simulteneous confidence intervals pairwise difference contrasts. But, in case of sets of more general contrasts, bonferroni's method works better. However, observe that the length of the confidence interval in scheffe's method does not increase even if more contrasts are added, while the length of CI given by bonferroni's method increases. This shows an indication that if there are lots of contrasts for which simulteneous confidence intervals are to be found, bonferroni's method will not remain better than Scheffe's method.

THANK YOU