Code

HW3-ADA-jq2292

Jing Qian (jq2282)

1.a. Compute 95% simultaneous confidence intervals for the differences in response between the four different diets using the Bonferroni method.

Hide

```
library(gdata)
data = read.xls("~/Desktop/4291-AdvancedDA/HW/HW2/HW2(Diets).xlsx", header=TRUE)
```

The 95% simultaneous confidence intervals for the differences in response between the four different diets using the Bonferroni method is:

$\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}} \pm t_{\frac{\gamma}{2g}}(n-k)\ SE(\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}})$.

Here $g = \binom{k}{2} = 6$, $t_{\frac{\gamma}{2g}}(n-k) = t_{\frac{0.05}{2*6}}(29-4) = t_{\frac{0.05}{12}}(25)$.

$SE(\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}}) = \sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$. And we could get MSE from the ANOVA table, or directly use the result of Homewok2, which is $MSE = 0.039096$.

So the result is calculated using the function within loops: $\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}} \pm t_{\frac{0.05}{12}}(25)\ \sqrt{0.039096}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$.

Hide

```
CI_Bonf = function(yi, yj, MSE, n, k, g){
  Lhat = mean(yi) - mean(yj)
  t = qt(1-0.05/(2*g), n-k)
  SEL = sqrt(MSE * (1/length(yi) + 1/length(yj)))
  dev = t  * SEL
  return(c(Lhat-dev, Lhat+dev))
}
for (i in 1:3){
  for (j in (i+1):4){
    cat("i =",i,", j =",j,", CI = (", CI_Bonf(data[,2][data[,1]==i], data[,2][data[,1]==
j],0.039096,29,4,6),")\n")
  }
}
```

```
i = 1 , j = 2 , CI = ( -0.1524595 0.4338881 )
i = 1 , j = 3 , CI = ( -0.1677714 0.4625333 )
i = 1 , j = 4 , CI = ( -0.4699595 0.1163881 )
i = 2 , j = 3 , CI = ( -0.2992597 0.3125931 )
i = 2 , j = 4 , CI = ( -0.6007328 -0.03426718 )
i = 3 , j = 4 , CI = ( -0.6300931 -0.01824026 )
```

1.b. Test for all pairs of factor level means whether or not they differ using the Scheffe’s procedure with α = 0.05. Set up groups of factor levels whose means do not differ.

$H_0 : L = \overline{Y_{i\cdot}} - \overline{Y_{j\cdot}} = 0.$

Using the Scheffe’s procedure, if $\frac{\hat{L}^2}{\widehat{Var(\hat{L})}} > (k-1)F_\gamma(k-1, n-k)$, we could reject H_0.

Here $\widehat{Var(\hat{L})} = MSE \sum_{i=1}^{k} \frac{c_i^2}{n_i} = MSE\,(\frac{1}{n_i} + \frac{1}{n_j})$, so we need to check whether

$\frac{(\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}})^2}{\frac{1}{n_i}+\frac{1}{n_j}} > (k-1) * MSE * F_\gamma(k-1, n-k) = (4-1) * 0.039096 * F_{0.05}(4-1, 29-4) = 0.3508367.$

Hide

```
val = 3*0.039096*qf(1-0.05, df1=3, df2=25)
cat("Critical value is:", val,"\n")
```

```
Critical value is: 0.3508367
```

Hide

```
for (i in 1:3){
  for (j in (i+1):4){
      yi = data[,2][data[,1]==i]
      yj = data[,2][data[,1]==j]
      Lhat = mean(yi) - mean(yj)
      ninv = 1/length(yi) + 1/length(yj)
      tmp = Lhat*Lhat/ninv
      if(tmp>val){
        s = "Reject H0\n"}
      else{
        s = "Accept H0\n"}
      cat("i =",i,", j =",j,", calculated value is: ",tmp,". ", s)
  }
}
```

```
i = 1 , j = 2 , calculated value is:  0.0739219 .  Accept H0
i = 1 , j = 3 , calculated value is:  0.07017601 .  Accept H0
i = 1 , j = 4 , calculated value is:  0.1166786 .  Accept H0
i = 2 , j = 3 , calculated value is:  0.000152381 .  Accept H0
i = 2 , j = 4 , calculated value is:  0.403225 .  Reject H0
i = 3 , j = 4 , calculated value is:  0.3602881 .  Reject H0
```

From above result, we could see that group 1 could not differ from group 2, 3, 4. Group 2 could not differ from group 1 and 3, but differs from 4. Group 3 could not differ from group 1 and 2, but differs from 4. The results we get using the Scheffe’s procedure agree with 1.(a), which is using the Bonferroni method. From the discussion above, we could not really set up groups of factor levels here.

2.a. Obtain the analysis of variance table and test whether or not the mean time lapse differs for the five agents.

Hide

```
data2 = read.table("~/Desktop/4291-AdvancedDA/HW/HW3/sofdrink.txt", header=TRUE)
```

Hide

```
anova(lm(data2$TimeLapse~factor(data2$Agent)))
```

```
Analysis of Variance Table

Response: data2$TimeLapse
                    Df Sum Sq Mean Sq F value    Pr(>F)
factor(data2$Agent)  4 4430.1 1107.53  147.23 < 2.2e-16 ***
Residuals           95  714.7    7.52
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Above is the ANOVA table. From the table, we could see that the p-value is less than 2.2e-16, which is far less than $\gamma = 0.05$. So we could reject the H_0 and so the mean time lapse differs for the five agents.

2.b. What is the least significant difference equal to when comparing agents 1 and 5? That is, what is the value of LSD_{15} when γ = 0.05.

$$LSD_{15} = t_{\frac{\gamma}{2}}(n-k)\sqrt{MSE}\sqrt{\frac{1}{n1}+\frac{1}{n5}} = t_{0.025}(95)\sqrt{7.52}\sqrt{\frac{1}{20}*2} = 1.721569.$$

Hide

```
t = qt(1-0.05/2, 95)
LSD15 = t*sqrt(7.52*2/20)
LSD15
```

```
[1] 1.721569
```

2.c. Test for all pairs of factor level means whether or not they differ using the Bonferroni procedure with γ = 0.05. Set up groups of factor levels whose means do not differ.

Hide

```
pairwise.t.test(data2$TimeLapse,data2$Agent,pool.sd=TRUE,p.adjust.method="bonf")
```

```
	Pairwise comparisons using t tests with pooled SD 

data:  data2$TimeLapse and data2$Agent 

  1       2       3       4      
2 0.2329  -       -       -      
3 < 2e-16 < 2e-16 -       -      
4 < 2e-16 3.1e-13 0.0067  -      
5 5.9e-08 9.7e-13 < 2e-16 < 2e-16

P value adjustment method: bonferroni 
```

From the table above, only $H_0 : \mu_1 - \mu_2 = 0$ has larger p-value than $\gamma = 0.05$ so we could not reject such H_0. Other pairs differ. So we could set up groups of factor levels as: {1, 2}, {3}, {4}, {5}.

2.d. Test for all pairs of factor level means whether or not they differ using the Scheffe's procedure with γ = 0.05. Set up groups of factor levels whose means do not differ.

$H_0 : L = \overline{Y_{i\cdot}} - \overline{Y_{j\cdot}} = 0$. Similar to question 1.b., for the Scheffe's procedure, if $\frac{(\overline{Y_{i\cdot}} - \overline{Y_{j\cdot}})^2}{\frac{1}{n_i} + \frac{1}{n_j}} > (k-1) * MSE * F_\gamma(k-1, n-k) = (5-1) * 7.52 * F_{0.05}(5-1, 100-5) = 74.22221$, we could reject H_0.

Hide

```
val2 = 4*7.52*qf(1-0.05, df1=4, df2=95)
cat("Critical value is:", val2,"\n")
```

Hide

```
for (i in 1:4){
  for (j in (i+1):5){
      yi = data2$TimeLapse[data2$Agent==i]
      yj = data2$TimeLapse[data2$Agent==j]
      Lhat = mean(yi) - mean(yj)
      ninv = 1/length(yi) + 1/length(yj)
      tmp = Lhat*Lhat/ninv
      if(tmp>val2){
        s = "Reject H0\n"}
      else{
        s = "Accept H0\n"}
      cat("i =",i,", j =",j,", calculated value is: ",tmp,". ", s)
  }
}
```

```
i = 1 , j = 2 , calculated value is:  40 .  Accept H0
i = 1 , j = 3 , calculated value is:  1638.4 .  Reject H0
i = 1 , j = 4 , calculated value is:  950.625 .  Reject H0
i = 1 , j = 5 , calculated value is:  308.025 .  Reject H0
i = 2 , j = 3 , calculated value is:  1166.4 .  Reject H0
i = 2 , j = 4 , calculated value is:  600.625 .  Reject H0
i = 2 , j = 5 , calculated value is:  570.025 .  Reject H0
i = 3 , j = 4 , calculated value is:  93.025 .  Reject H0
i = 3 , j = 5 , calculated value is:  3367.225 .  Reject H0
i = 4 , j = 5 , calculated value is:  2340.9 .  Reject H0
```

From the result above, we could set up following groups of factor levels: {1, 2}, {3}, {4}, {5}.

2.e. The marketing director wishes to compare the mean time lapses for agents 1, 3 and 5. Obtain the pairwise confidence interval for all pairwise comparisons among these three treatment means using the Bonferroni procedure with a 90% family confidence coefficient. Interpret your result.

Hide

```
CI_Bonf2 = function(yi, yj, MSE, n, k, g){
  Lhat = mean(yi) - mean(yj)
  t = qt(1-0.1/(2*g), n-k)
  SEL = sqrt(MSE * (1/length(yi) + 1/length(yj)))
  dev = t  * SEL
  return(c(Lhat-dev, Lhat+dev))
}
cat("i =",1,", j =",3,", CI = (", CI_Bonf2(data2$TimeLapse[data2$Agent==1], data2$TimeLa
pse[data2$Agent==3],7.52,100,5,3),")\n")
```

```
i = 1 , j = 3 , CI = ( 10.92737 14.67263 )
```

Hide

```
cat("i =",1,", j =",5,", CI = (", CI_Bonf2(data2$TimeLapse[data2$Agent==1], data2$TimeLa
pse[data2$Agent==5],7.52,100,5,3),")\n")
```

```
i = 1 , j = 5 , CI = ( -7.422629 -3.677371 )
```

Hide

```
cat("i =",3,", j =",5,", CI = (", CI_Bonf2(data2$TimeLapse[data2$Agent==3], data2$TimeLa
pse[data2$Agent==5],7.52,100,5,3),")\n")
```

```
i = 3 , j = 5 , CI = ( -20.22263 -16.47737 )
```

From the pariwise comparisons among agents 1, 3 and 5, we could find that none of the confidence intervals with 90% family confidence intervals covers 0. $H_0 : \mu_1 - \mu_3 = 0, H_0 : \mu_1 - \mu_5 = 0$ and $H_0 : \mu_3 - \mu_5 = 0$ are all rejected, which means each agent belongs to a different group of factor levels.

2.f. Agents 1 and 2 distribute merchandise only, agents 3 and 4 distribute cash-value coupons only and agent 5 distributes both merchandise and coupons. Estimate the contrast $L = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$ using a 95% confidence interval. Interpret your result.

Hide

```
library(gmodels)
fit.contrast(lm(data2$TimeLapse~factor(data2$Agent)), factor(data2$Agent), c(0.5,0.5,-0.
5,-0.5,0), conf=0.95)
```

```
                                              Estimate Std. Error  t value     Pr(>|t|)
lower CI upper CI
factor(data2$Agent) c=( 0.5 0.5 -0.5 -0.5 0 )   10.275  0.6132957 16.75375 4.086083e-30
9.057454 11.49255
attr(,"class")
[1] "fit_contrast"
```

We could see that the estimated 95% confidence interval is (9.057454, 11.49255), which doesn't include 0. That means we could reject H_0. So the mean of agents 1 and 2 differs from that of agents 3 and 4. In other words, the two ways of distributions: merchandise only and cash-value coupons only have different time lapse.

2.g. Estimate the following comparisons with a 95% confidence interval using the Bonferroni method $L_1 = \mu_1 - \mu_2,\ L_2 = \frac{\mu_1+\mu_2}{2} - \mu_5,\ L_3 = \frac{\mu_3+\mu_4}{2} - \mu_5$.

Using the self-defined function of 95% confidence interval using the Bonferroni method, we could calculate that for each L. Here the time of comparisons is 3. MSE is 7.52, from the ANOVA table in 2.(a). $n = 100,\ k = 5$. So we have:

Hide

```
g = 3
CI1 = CI_Bonf(data2$TimeLapse[data2$Agent==1], data2$TimeLapse[data2$Agent==2], 7.52, 10
0, 5, g)
t = qt(1-0.05/(2*g), 100-5)
MSE = 7.52
SEL2 = sqrt(MSE * (0.5*0.5*2+1)/20)
dev = t * SEL2
Lhat2 = mean((data2$TimeLapse[data2$Agent==1]+data2$TimeLapse[data2$Agent==2])/2) - mean
(data2$TimeLapse[data2$Agent==5])
CI2 = c(Lhat2 - dev, Lhat2 + dev)
Lhat3 = mean((data2$TimeLapse[data2$Agent==3]+data2$TimeLapse[data2$Agent==4])/2) - mean
(data2$TimeLapse[data2$Agent==5])
CI3 = c(Lhat3 - dev, Lhat3 + dev)
cat("The 95 percent confidence interval for L1 is: (", CI1,").\n")
```

```
The 95 percent confidence interval for L1 is: ( -0.1134118 4.113412 ).
```

Hide

```
cat("The 95 percent confidence interval for L2 is: (", CI2,").\n")
```

```
The 95 percent confidence interval for L2 is: ( -8.380268 -4.719732 ).
```

Hide

```
cat("The 95 percent confidence interval for L3 is: (", CI3,").\n")
```

```
The 95 percent confidence interval for L3 is: ( -18.65527 -14.99473 ).
```

From the calculated confidence interval, we could find that agent 1 and agent 2 could not differ. The mean of agent 1 and agent 2 differ from agent 5. The mean of agent 3 and agent 4 differ from agent 5.