5 Multiple Comparisons

1. The following code constructs a vector of normal random numbers, y and a factor named group, and then displays the data with boxplots. Test whether the groups have equal variance.

```r
set.seed(98081)
y <- c(rnorm(n=10,mean=0,sd=3),rnorm(10,1,3),rnorm(10,2,3),rnorm(10,4,3))
group <- gl(4,10,labels=c("g1","g2","g3","g4"))
boxplot(y~group)  # use this to inspect data
```

![Boxplot of groups g1, g2, g3, g4](image)

```r
bartlett.test(y~group)  # evaluate constant variance assumption
```

```r
## Bartlett test of homogeneity of variances
##
c10.00000000000000  c20.00000000000000  c30.00000000000000  c40.00000000000000
c
```
## data: y by group
## Bartlett's K-squared = 4.639, df = 3, p-value = 0.2

2. Use `aov` to evaluate the null hypothesis of no difference among group means.

```r
y.aov.01 <- aov(y~group)  # anova
summary(y.aov.01)
```

```
# Df Sum Sq Mean Sq F value Pr(>F)
group 3 110 36.5 4.66 0.0075 **
Residuals 36 282 7.8
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Use the command `?TukeyHSD` to read the help page for that function. Use `TukeyHSD` to evaluate all pairwise differences among the groups while maintaining a family-wise Type I error rate of 0.05.

```r
TukeyHSD(y.aov.01)  # Tukey HSD
```

```
# Tukey multiple comparisons of means
# 95% family-wise confidence level

# Fit: aov(formula = y ~ group)
#
# $group
diff  lwr  upr p adj
g2-g1 0.515446 -2.8549392 3.88583 0.976074
g3-g1 0.794261 -2.5761236 4.16465 0.920131
g4-g1 4.201316 0.8309310 7.57170 0.009664
g3-g2 0.278816 -3.0915692 3.64920 0.996026
g4-g2 3.685870 0.3154855 7.05626 0.027529
g4-g3 3.407055 0.0366698 6.77744 0.046739
```

4. Next, evaluate all pairwise differences among the groups while maintaining a family-wise Type I error rate of 0.01. How do the results differ from the ones obtained in the previous question?

```r
TukeyHSD(y.aov.01,conf.level=0.99)  # wider CIs but same p values
```

```
# Tukey multiple comparisons of means
# 99% family-wise confidence level

# Fit: aov(formula = y ~ group)
#
# $group
diff  lwr  upr p adj
g2-g1 0.515446 -3.6696150 4.70051 0.976074
g3-g1 0.794261 -3.3907994 4.97932 0.920131
g4-g1 4.201316 0.0162553 8.38638 0.009664
```
## g3-g2 0.278816 -3.9062449 4.46388 0.996026
## g4-g2 3.685870 -0.4991903 7.87093 0.027529
## g4-g3 3.407055 -0.7780059 7.59212 0.046739

Table 1: Levels of the variable `hostbird` in the dataframe `cuckoo`.

<table>
<thead>
<tr>
<th>level</th>
<th>bird</th>
</tr>
</thead>
<tbody>
<tr>
<td>hdgesprw</td>
<td>Hedge Sparrow</td>
</tr>
<tr>
<td>mdwpipit</td>
<td>Meadow Pipit</td>
</tr>
<tr>
<td>piedwtail</td>
<td>Pied Wagtail</td>
</tr>
<tr>
<td>robin</td>
<td>Robin</td>
</tr>
<tr>
<td>treepipit</td>
<td>Tree Pipit</td>
</tr>
<tr>
<td>wren</td>
<td>Wren</td>
</tr>
</tbody>
</table>

5. Read the cuckoo data file (described in Lab #4) into R. The data frame contains the variable `egglength`, which is the length of a cuckoo’s egg (in mm) and the factor `hostbird`, which is the name of the host bird (i.e., the foster parent) of each egg. The levels of `hostbird` are shown in Table 1.

```r
cuckoo <- read.csv("http://psycserv.mcmaster.ca/bennett/psy710/datasets/cuckoo.csv")
```

(a) Calculate the group means and n.

```r
(egg.m <- with(cuckoo,tapply(egglength,hostbird,mean)) ) # group means
## hdgesprw mdwpipit piedwtail robin treepipit wren
## 23.1214 22.2989 22.9033 22.5750 23.0900 21.1300
(egg.n <- with(cuckoo,tapply(egglength,hostbird,length)) ) # unequal group n
## hdgesprw mdwpipit piedwtail robin treepipit wren
## 14 45 15 16 15 15
```

(b) Examine the data to determine if the groups have equal variance.

```r
boxplot(egglength~hostbird,data=cuckoo); # graphical test
```
Bartlett's test of homogeneity of variances

Bartlett's K-squared = 4.479, df = 5, p-value = 0.483

(c) The Tukey HSD procedure assumes there are equal n per group, and that the group variances are equal. We do not have equal n per group, and so the Tukey HSD procedure is not appropriate here. However, the Tukey-Kramer test does not assume equal n. The Tukey-Kramer test can be performed with the `TK.test` function in the `DTK` package. Install and load the `DTK` package into R with the following commands and then use it to evaluate all pairwise differences between groups. Set the family-wise Type I error rate to 0.05.

```
#install.packages("DTK") # download package and install on computer
library("DTK") # load package into workspace

TK.test(x=cuckoo$egglength,f=cuckoo$hostbird,a=0.05) # Tukey-Kramer test
```

```
# Tukey multiple comparisons of means
# 95% family-wise confidence level
#
# Fit: aov(formula = x ~ f)
#
# $f
# diff lwr upr p adj
# mdwpipit-hdgesprw -0.8225397 -1.62913361 -0.0159458 0.042862
# piedwtail-hdgesprw -0.2180952 -1.19755944 0.7613690 0.987219
# robin-hdgesprw -0.5464286 -1.51100320 0.4181461 0.572615
```
Answer: The Tukey-Kramer test shows that all of the pairwise tests including wrens are significant: egg length in each host bird differs from egg length in wrens. Egg length in Meadow Pipits also differed from egg length in Hedge Sparrows ($p_{adj} = 0.042$) and Tree Pipits ($p_{adj} = 0.047$).

(d) The following code illustrates how to use the `subset` function to extract the part of the cuckoo data frame that corresponds to the `hdgesprw` group, and store the result in a data frame named `hedgesparrow`:

```r
levels(cuckoo$hostbird)
hedgesparrow <- subset(cuckoo, hostbird == "hdgesprw")
```

Use `subset` to extract the data for Hedge Sparrows, Meadow Pipits, Robins, and Wrens. Store each subset of data in separate variables named `hedgesparrow`, `meadowpippit`, `robin`, and `wren`. Inspect the variables to make sure that they contain data only from birds of the correct type.
(e) Use \( t \) tests (assuming equal group variances) to evaluate the null hypothesis of no difference between average egg lengths in i) Hedge Sparrows and Robins; ii) Hedge Sparrows and Meadow Pipits; and iii) Wrens and Robins.

\[
\text{hedgesparrow} \leftarrow \text{subset(cuckoo, hostbird=="hdgesprw")}
\]
\[
\text{meadowpippit} \leftarrow \text{subset(cuckoo, hostbird=="mdwpipit")}
\]
\[
\text{robin} \leftarrow \text{subset(cuckoo, hostbird=="robin")}
\]
\[
\text{wren} \leftarrow \text{subset(cuckoo, hostbird=="wren")}
\]

\[
\text{t.test(hedgesparrow$egglength,robin$egglength,var.equal=TRUE)}
\]

##
## Two Sample t-test
##
data:  hedgesparrow$egglength and robin$egglength
## t = 1.689, df = 28, p-value = 0.102
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.11622  1.20908
## sample estimates:
## mean of x mean of y
## 23.1214  22.5750

\[
\text{t.test(hedgesparrow$egglength,meadowpippit$egglength,var.equal=TRUE)}
\]

##
## Two Sample t-test
##
data:  hedgesparrow$egglength and meadowpippit$egglength
## t = 2.81, df = 57, p-value = 0.00677
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.236438  1.408642
## sample estimates:
## mean of x mean of y
## 23.1214  22.2989

\[
\text{t.test(robin$egglength,wren$egglength,var.equal=TRUE)}
\]

##
## Two Sample t-test
##
data:  robin$egglength and wren$egglength
## t = 5.633, df = 29, p-value = 4.38e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.920353  1.969647
## sample estimates:
## mean of x mean of y
## 22.575  21.130

(f) Assuming that you used a per-comparison Type I error rate of .05, what is the family-wise Type I error rate for all 3 \( t \) tests?
alpha.pc <- .05;
C <- 3;
(alpha.fw <- 1 - (1-alpha.pc)^C)
## [1] 0.142625

(g) Assuming that the three comparisons are planned, how should you adjust your analyses to control
the family-wise Type I error rate or the False Discovery Rate?

# Bonferroni adjustment
alpha.fw <- .05; # desired fw Type-I error rate
C <- 3; # number of tests
(alpha.pc <- alpha.fw/C )
## [1] 0.0166667

# using p.adjust:
p.values <- c(0.102,4.38e-6,0.00677)
p.adjust(p.values,'bonferroni') # too conservative?
## [1] 3.060e-01 1.314e-05 2.031e-02

p.adjust(p.values,'holm') # more powerful
## [1] 1.020e-01 1.314e-05 1.354e-02

p.adjust(p.values,'fdr') # False-Discovery Rate=.05 (most powerful)
## [1] 0.10200000 0.00001314 0.01015500

(h) After making that adjustment, are your t tests significant?

**Answer:** Let’s assume that we set our family-wise error rate or false discovery rate to 0.05 (i.e.,
$\alpha_{FW} = 0.05$ or FDR = 0.05). If an adjusted $p$ value is less than or equal to $\alpha_{FW}$ (or FDR) than that
statistical test is significant. In this case, the first $t$ test (between Hedge Sparrows and Robins) is
not significant, but the other $t$ tests are statistically significant and we reject the null hypothesis of
no difference for both of those comparisons.

6. Conduct three linear contrasts (i.e., not $t$ tests) that evaluate the hypothesis of no difference in egg length
between i) Hedge Sparrows vs. Meadow Pipits; ii) (Hedge Sparrows & Pied Wagtails & Tree Pipits) vs.
(Meadow Pipits & Robins); and iii) Robins vs. Tree Pipits. Assume that the contrasts are planned and
adjust your procedures to maintain a family-wise Type I error rate of .05.

**Answer:** The following code performs the linear contrasts:

```r
source(url("http://psycserv.mcmaster.ca/bennett/psy710/Rscripts/linear_contrast_v2.R"))
```

```r
## [1] "loading function linear.comparison"

# Bonferroni adjustment:
(p.critical <- .05/3) # p-values less than this are significant

## [1] 0.0166667

levels(cuckoo$hostbird)
c1 <- c(1,-1,0,0,0,0)  # hdgesprw vs mdwpipit

c2 <- c(1/3, -1/2, 1/3, -1/2, 1/3, 0)  # (hdgesprw & piedwtail & treepipit) vs. (mdwpipit & robin)

c3 <- c(0,0,0,1,-1,0);  # robin vs treepipit

myContrasts <- list(c1,c2,c3);

tmp <- linear.comparison(y=cuckoo$egglength,group=cuckoo$hostbird,c.weights=myContrasts)

# computing linear comparisons assuming equal variances among groups

# compare p values to p.critical
# or compute adjusted p-values (significant if p.adj < .05):

p.adjust(p=c(.004,.002,.118),method='bonferroni')

# orthogonal if sum of cross-products is zero:

sum(c1*c2)  # not orthogonal

sum(c1*c3)  # orthogonal

sum(c2*c3)  # not orthogonal

# Are the results of the first contrast (which compared egg lengths in Hedge Sparrows and Meadow Pipits) the same as the t test performed from a previous question that compared these two groups? Why or why not?

Answer: No, the two tests yield different values of t and p. The t statistic expresses the difference between group means relative to an estimate of the population error variance. The two procedures yield different results because the linear contrast uses the entire data set to estimate the error (and the degrees of freedom) whereas the t test uses only the values in the two groups being compared.
9. Now assume that the decision to do the three linear comparisons in the previous questions was taken after looking at the data. How could you adjust your analyses to maintain a family-wise Type I error rate of .05? Are your comparisons still significant after making this adjustment?

```r
# Scheffe test:
alpha <- .05; # desired Type I error rate
a <- length(egg.m) # number of different groups
N <- sum(egg.n) # total N
F.critical <- qf(1-alpha, df1=a-1, df2=N-a) # critical F for omnibus test
(F.scheffe <- (a-1)*F.critical) # critical Scheffe F value

## [1] 11.4696

F1 <- 8.738
F2 <- 9.955
F3 <- 2.484
(F1 > F.scheffe) # 1st comparison is NOT significant

## [1] FALSE

(F2 > F.scheffe) # 2nd comparison is NOT significant

## [1] FALSE

(F3 > F.scheffe) # 3rd comparison is NOT significant

## [1] FALSE
```