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
```

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

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.

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?
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 $n$.

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

(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.

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

(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`, `meadowpipit`, `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.

(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?

(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?

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

7. Are these three contrasts orthogonal?

8. 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?

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?