Statistics Lab #5
Multiple Comparisons

5 Multiple Comparisons

5.1 Initialize R

Before starting this lab, please initialize R by entering the following commands at the prompt:

```r
> options(contrasts = c("contr.sum", "contr.poly")) # sum-to-zero definition of effects
> options(digits = 6, width = 70)
```

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>hedgesprw</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>

5.2 Exercises

1. The following code constructs a vector of normal random numbers, y, and a factor named group:

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

Test whether the groups have equal variance. Then use a procedure to evaluate all pairwise differences among the groups while maintaining a familywise Type I error rate of 0.05.

2. Read the cuckoo data file (described in Lab #3) into R with the following command:

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

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. Examine the data to determine if the groups have equal n and/or constant variance. Then evaluate each pairwise difference between groups, testing the null hypothesis that the difference between means is zero. Set the familywise Type I error rate to 0.05.
3. 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
> hedgesparrow <- subset(cuckoo, hostbird == "hdgesprw")
```

Use `subset` to extract the data for Hedge Sparrows, Meadow Pippits, 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.

4. The following code illustrates the use of `t.test` to compare the mean egg length in Hedge Sparrows and Robins, assuming the two groups have equal variance:

```r
> t.test(hedgesparrow$egglength, robin$egglength, var.equal = TRUE)
```

(a) Use `t` tests (assuming equal group variances) to evaluate the null hypotheses of no difference between average egg lengths in i) Hedge Sparrows and Meadow Pippits; and ii) Wrens and Robins.

(b) Assuming that you used a per-comparison Type I error rate of .05, what is the family-wise Type I error rate?

(c) Assuming that the comparisons are planned, how should you adjust your analyses so that your family-wise error rate is .05?

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

5. The following code calculates the mean and `n` for each group:

```r
> egg.m <- with(cuckoo, tapply(egglength, hostbird, mean)) # group means
> egg.n <- with(cuckoo, tapply(egglength, hostbird, length)) # group n
```

Use this information to conduct two linear contrasts (i.e., not `t` tests) that evaluate the hypothesis of no difference in egg length between i) Hedge Sparrows and Meadow Pippits; and ii) Wrens and Robins. Assume that the contrasts are planned, and adjust your procedures to maintain a family-wise Type I error rate of .05.

6. Now assume that the decision to do the two pairwise 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?