STAT502 Lab #3

In this Lab we will use R to do multiple comparisons, which are usually considered after the original ANOVA F-test is found significant.

The data set comes from a study on three different drugs and the reduction of systolic blood pressure over one month. Denote by μ_1, μ_2, μ_3 the population mean systolic blood pressure reductions for the three drug groups. The data set is available on Canvas and can be read into R with the following commands:

```
data = read.table('bp.dat',header=T)
y = data[,1] # these are the blood pressure reductions
x = as.factor(data[,2]) # these are the drug group labels
```

1. Construct side-by-side boxplots for the data, and comment on the evidence for different group means.

2. Use the ANOVA model to carry out the test for equal μ_i. Report the test statistic, degrees of freedom, p-value, and conclusion with $\alpha = .05$.

3. If the null hypothesis wasn't rejected above, we could stop now. Alas, we all know it was rejected, and so the question remains: "which means are different?" For pairwise comparisons (group i versus group k), recall the test statistic is

$$\frac{\overline{Y}_{i\cdot} - \overline{Y}_{k\cdot}}{\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_k})}}$$

4. If we were interested in only a single comparison, say $\mu_1 - \mu_2$, we would use the usual upper $\alpha/2$ critical value from the t-distribution with $n_T - r$ degrees of freedom. What would the critical value be in this case?

5. If, however, we are interested in *all* pairwise comparisons (1vs2, 1vs3, and 2vs3), then we have to control for the "family-wise Type I error rate" (FWER). There is more than one way to do this.

Tukey's method:

Recall the Tukey critical value based on the studentized range distribution:

$$T = \frac{1}{\sqrt{2}} q_{r,n_T-r}(\alpha)$$

(a) Use the R function `qtukey()` to find the q value, and compute the value of T from this. How does it compare with value from 4 above?

(b) If we use the Tukey value instead of the value in 4, are we more or less likely to reject $H_0 : \mu_i - \mu_k = 0$? Explain why this makes sense if we are trying to control the FWER.

(c) Use the Tukey value T, along with the test statistic in 3 to test each pair of means. Which are significantly different? The following code may be helpful:

```
ybar = tapply(y,x,mean)
n = tapply(y,x,length)
i=1; k=2
se = sqrt(mse*(1/n[i]+1/n[k]))
test.stat = (ybar[i]-ybar[k])/se
```

(d) R also has a function that can effectively the same thing: `TukeyHSD(aov(y~x))`. The output gives the confidence intervals and p-values for each pairwise comparison, adjusted to control the FWER rate.

(e) Finally, the command `plot(TukeyHSD(aov(y~x)))` plots the confidence limits with a (dotted) reference line at 0. What role does this line play?

Bonerroni's method:

Recall the Bonferroni approach is to divide the desired FWER by the number of tests g. Specifically, if the desired FWER is α, then the Bonferroni value is $\alpha' = \alpha/g$.

$$B = t_{n_T - r}(\alpha'/2)$$

(f) Use the `qt()` function with the adjusted α' to find the Bonferroni critical value. How does it compare with the values in 4 and 5(a) above?

(g) If both the Tukey and Bonferroni critical values control the FWER, which would we prefer to use in this situation? Why?

(h) Compare all pairs of means using the Bonferroni critical value. Which are significant? Do the results agree with those in 5(c)?

Scheffé's method:

Consider the family of linear combinations of the form $L = \sum_{i=1}^{r} c_i \mu_i$ for some given constants $c_1, \ldots, c_r$. Tests of interest here are of the form $H_0 : L = 0$, and the test statistic is

$$\frac{\sum_{i=1}^{r} c_i \overline{Y}_{i\cdot}}{\sqrt{MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}}}$$

Note that all the pairwise comparisons $\mu_i - \mu_k$ are special cases of this general form. Scheffé's critical value for testing hypotheses of this form is

$$S = \sqrt{(r-1) f_{r-1, n_T - r}(\alpha)}$$

where $f_{r-1, n_T - r}(\alpha)$ is the F critical value used in the original ANOVA test.

(i) What is the value of S in this situation? How does it compare with the other critical values above?

(j) Compare all pairs of means using the Scheffé critical value. Which are significant? Do the results agree with those in 5(c)?

(k) For each of the three approaches considered here, Tukey, Bonferroni, and Scheffé, give one advantage of each.