Multiple Comparisons

1. Objectives
- Determine when to use multiple comparison tests.
- Conduct multiple comparison tests with Bonferroni and Tukey's HSD method.
- Produce and interpret the results.

2. Exercises

Firstly, remember to set your own working directory.

Exercise 1. Kudzu is a plant that was imported to the United States from Japan and now covers over seven million acres in the South. The plant contains chemicals called *isoflavones* that have been shown to have beneficial effects on bones. One study used three groups of rats to compare a control group with rats that were fed either a low dose or a high dose of isoflavones from **kudzu.csv**. One of the outcomes examined was the *bone mineral density* in the femur (in grams per square centimeter).

    a. Use graphical and numerical methods to describe the data.
    b. Examine the assumptions necessary for one-way ANOVA. Summarize your findings.
    c. Run one-way ANOVA and report the results.
    d. Use a multiple-comparisons method to compare the three groups.
    e. Write a short report explaining the effect of kudzu isoflavones on the femur of the rat.

Let's run the following codes:

```
➢ kudzu<-read.table("kudzu.csv", header=TRUE, sep = ",",
  stringsAsFactors = F)
➢ str(kudzu) #to see the structure of the data frame
➢ kudzu$treatment<-factor(kudzu$treatment, levels =
  c("Control","LowDose","HighDose"))
➢ kudzu$treatment
➢ table(kudzu$treatment) #check the sample sizes
```

Just follow the instructions from **Lab 4** for part a and b. In this lab, we focus on multiple comparison tests, and so we must conduct one-way ANOVA first to see whether significant difference exists among 3 groups.

```
➢ aovKudzu <- aov(boneMineralDensity~treatment,data=kudzu)
➢ summary(aovKudzu)
```

You should get the following output:

```
            Df   Sum Sq   Mean Sq F value Pr(>F)
treatment    2 0.003186 0.0015928   7.718 0.0014 **
Residuals   42 0.008668 0.0002064
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on this result, we can conclude that there exists a significant difference among 3 groups of Control, LowDose, and HighDose. Then we can apply a multiple comparison method to identify where this difference comes from.

For Bonferroni method, we will use the function **pairwise.t.test()** for comparing all pairs of means simultaneously. The format of the code is as the following:

➢ `pairwise.t.test(outcome, grouping factor, p.adjust.method = p.adjust.methods, paired = FALSE)`

Below are the meanings of the arguments:

`outcome`	Name of your outcome variable
`grouping factor`	Name of the factor that defines the groups
`p.adjust.method`	Method for adjusting p-value. Use **"bonferroni"** if you want Bonferroni method.
`paired`	A logical statement indicating whether you want paired-samples t-tests

So the code to run the Bonferroni method for our exercise will be:

➢ `pairwise.t.test(kudzu$boneMineralDensity, kudzu$treatment, paired=F, p.adjust.method = "bonferroni")`

Here is the ouput:

```
        Pairwise comparisons using t tests with pooled SD
data:  kudzu$boneMineralDensity and kudzu$treatment
         Control LowDose
LowDose  1.0000  -
HighDose 0.0107  0.0022

P value adjustment method: bonferroni
```

For Tukey's HSD method, we need the **multcomp** package. First, install the package. Then use **library(multcomp)** to load the package. The appropriate function is **glht()**, the format of which is as follows:

➢ `mcpModel <- glht(aovModel, linfct = mcp(grouping factor="method"))`
➢ `summary(mcpModel)`

where:

`mcpModel`	Object containing information of the multiple comparison procedure used
`aovModel`	Name of the model that has been created with the `aov()` function.
`grouping factor`	Factor that defines the groups
`linfct = mcp(grouping factor="method")`	Specifies the method to adjust p-value. If you want **Tukey's HSD**, replace "method" by "Tukey"
`summary(mcpModel)`	Print output of the multiple comparison procedure contained in mcpModel

So the code to run the Tukey's HSD method for our exercise will be:

- ➢ `library(multcomp)`
- ➢ `tukeyKudzu <- glht(aovKudzu, linfct = mcp(treatment = "Tukey"))`
- ➢ `summary(tukeyKudzu)`

You will see the following output:

```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = boneMineralDensity ~ treatment, data = kudzu)

Linear Hypotheses:
                       Estimate Std. Error t value Pr(>|t|)
LowDose - Control == 0  -0.002933   0.005246  -0.559  0.84234
HighDose - Control == 0  0.016200   0.005246   3.088  0.00961 **
HighDose - LowDose == 0  0.019133   0.005246   3.648  0.00200 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

Question 1. In which case should we choose Bonferroni or Tukey's HSD method? Which method is more appropriate for this problem?

Question 2. Read and interpret the outputs for Bonferroni and Tukey's HSD methods. Compare the 2 outputs and make comments about their differences.

Exercise 2. If a supermarket product is frequently offered at a reduced price, do customers expect the price of the product to be lower in the future? This question was examined by researchers in a study conducted on students enrolled in an introductory management course at a large midwestern university. For 10 weeks, 160 subjects read weekly ads for the same product. Students were randomly assigned to read 1, 3, 5, or 7 ads featuring price promotions during the 10-week period. They were then asked to estimate what the product's price would be the following week.

(a)   Make a Q-Q plot for the data in each of the four treatment groups. Summarize the information in the plots and draw a conclusion regarding the **normality** of data.
(b)   Summarize the data with a table containing the sample size, mean, and standard deviation for each group.
(c)   Is the assumption of equal standard deviations reasonable here? Explain why or why not.
(d)   Carry out a one-way ANOVA. Give the hypotheses, the test statistic with its degrees of freedom, and the p-value. State your conclusion.
(e)   Use the Bonferroni or Tukey's HSD procedure to compare the group means. Summarize the results and support your conclusions with a graph of the means.

Part a, b, c are your homework (follow the instructions from Lab 4). After importing data from **pricepromo.csv**, you're expected to produce output for d and e as follows:

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Promotions   3  8.361  2.7868   25.66 1.52e-13 ***
Residuals  156 16.946  0.1086
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


        Pairwise comparisons using t tests with pooled SD

data:  price$Price and price$Promotions

     1ads     3ads     5ads
3ads 0.2695   -        -
5ads 4.5e-07  0.0023   -
7ads 1.7e-12  9.6e-08  0.1242

P value adjustment method: bonferroni


        Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = Price ~ Promotions, data = price)

Linear Hypotheses:
                  Estimate Std. Error t value Pr(>|t|)
3ads - 1ads == 0  -0.1490     0.0737  -2.022  0.18438
5ads - 1ads == 0  -0.4163     0.0737  -5.648  < 0.001 ***
7ads - 1ads == 0  -0.5885     0.0737  -7.985  < 0.001 ***
5ads - 3ads == 0  -0.2672     0.0737  -3.626  0.00204 **
7ads - 3ads == 0  -0.4395     0.0737  -5.963  < 0.001 ***
7ads - 5ads == 0  -0.1722     0.0737  -2.337  0.09417 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```