Lab 7

Basic Analysis of Variance in R
Psychology 310

Instructions. Work through the lab, saving the output as you go. If you work in Microsoft Word, you can easily copy any graph to Word via the clipboard. Numerical output may also be copied easily by highlighting, moving it to the clipboard, then copying into Word. However, you should format R output in TrueType Courier New font so that it is *monospaced*. Your output file should be named `LAST_FIRST_LAB7.DOC`, where `LAST` is your last name, and `FIRST` is your first name. Any additional files should have the same naming scheme, except the file extension should be correct. You may add any description text you wish after `LAB6` in the file name.

Preamble. Today's assignment involves the analysis of variance.

1 Introduction

The analysis of variance (often abbreviated ANOVA or AOV) is a statistical technique that has long played an important role in psychological research. Most of you will go on to take Psychology 304B, which spends a significant amount of time on ANOVA. This is just a brief introduction to performing ANOVA analyses with R.

2 One Way Fixed-Effects ANOVA

In the one way, fixed-effects ANOVA, we test the hypothesis that J groups have equal means. With the fixed-effects model, we are only allowed to make inferences about the populations actually sampled in the design. So, for example, if we have 3 groups, one of them is a control group, one of the groups receives 5 units of a drug, and one receives 10 units, strictly speaking we can only make inferences about those 3 situations.

You textbook discusses the "linear model" for ANOVA, and the ANOVA source table. R has several libraries that handle ANOVA calculations, from very simple designs to extremely complex ones. We'll just look at how to perform calculations for the most basic designs here.

2.1 Data Setup

Problem 5, page 417 in Glass and Hopkins, discusses a simple experiment with $J = 3$ groups, and the following data:

Group I Postorganizer	Group II Preorganizer	Group III NoOrganizer
5	4	5
4	5	4
4	3	6
7	6	2
8	6	2
7	3	2
6	3	6
4	4	4
4	4	3
7	2	5

To enter these data in a form suitable for analysis by R, we need to create two variables, one of which is a `factor` variable, the other a `numeric` variable. There are many ways we can do this. One way is to create the data frame like this: First, enter the group labels, using the R `rep` function to help save you time and effort:

```
> group <- c(rep("Postorganizer",10),rep("Preorganizer",10),rep("NoOrganizer",10))
```

Next, enter the scores as one long vector:

```
> score <- c(5,4,4,7,8,7,6,4,4,7,4,5,3,6,6,3,3,4,4,2,5,4,6,2,2,2,6,4,3,5)
```

We want to make sure that our ANOVA procedures process the `group` variable as a special variable type, called `factor`. One way to do this is to convert the variable, as follows:

```
> group <- factor(group)
```

We can verify that `group` is now a factor variable.

```
> is.factor(group)
[1] TRUE
```

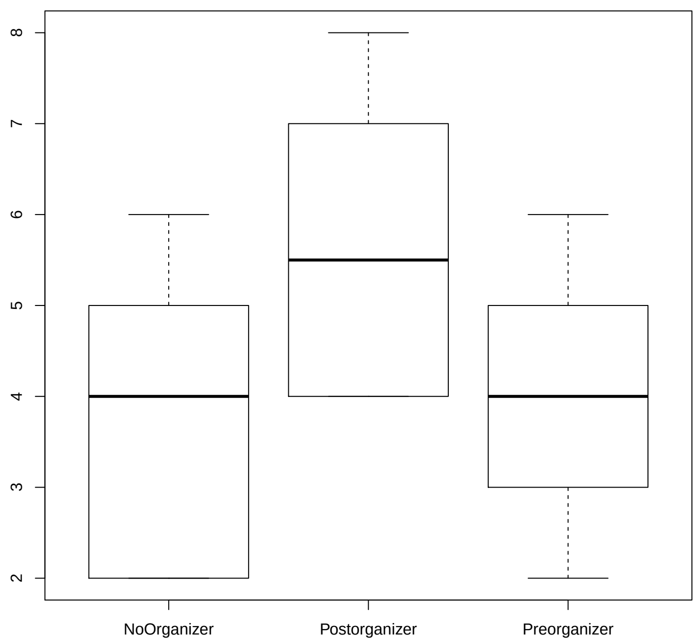
We can, if we want, create a data frame and save it. For example:

```
> GH417.data <- data.frame(group,score)
> write.table(GH417.data,"GH417.txt",col.names=TRUE,row.names=FALSE)
```

2.2 ANOVA computation

Before performing the ANOVA, we can do some preliminary exploratory analyses by doing a boxplot.

```
> boxplot(score~group)
```



To perform the one way, fixed-effects ANOVA on these data, we can use either the `lm()` or `aov()` function. In fact, `aov` calls `lm` to do its computations. To perform the ANOVA, we create an ANOVA fit object as

```
> fit <- aov(score ~ group)
```

The results can be viewed as

```
> summary(fit)
            Df Sum Sq Mean Sq F value  Pr(>F)
group        2   18.2  9.1000  4.0082 0.02991 *
Residuals   27   61.3  2.2704
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compare these to the answer key in Glass and Hopkins, page 421.

2.3 Tukey Pairwise Contrasts

The Tukey pairwise contrast procedure discussed in the book can be performed as follows.

First, create a TukeyHSD object by operating on the ANOVA fit object

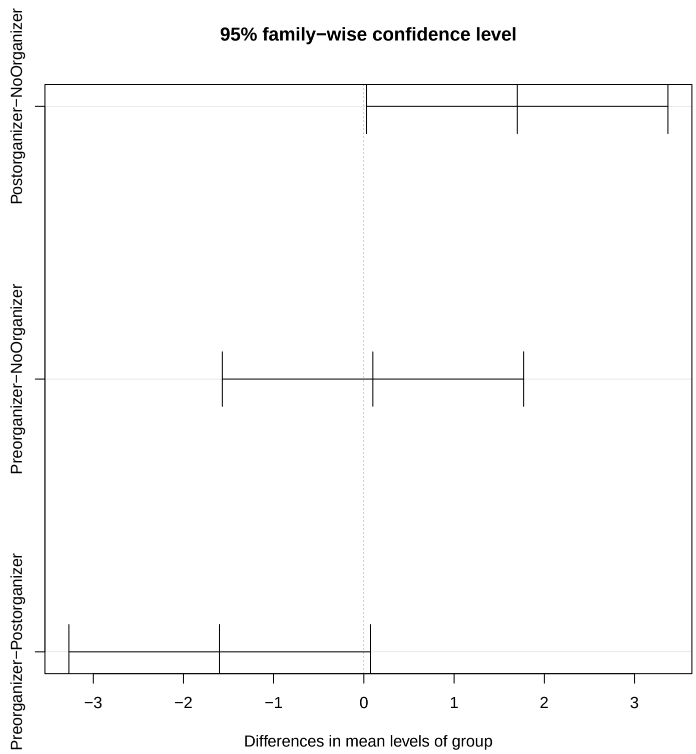
```
> TukeyTestResults <- TukeyHSD(fit,"group")
> TukeyTestResults
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = score ~ group)

$group
                           diff         lwr        upr     p adj
Postorganizer-NoOrganizer   1.7  0.02924393 3.37075607 0.0455236
Preorganizer-NoOrganizer    0.1 -1.57075607 1.77075607 0.9879376
Preorganizer-Postorganizer -1.6 -3.27075607 0.07075607 0.0624878
```

You can see from the results that the Postorganizer and NoOrganizer groups are significantly different, as the confidence interval excludes zero. You can plot the confidence intervals as follows:

```
> plot(TukeyTestResults)
```



3 On Your Own

The file *hsb.txt* has the High School and Beyond data set distributed with the Glass and Hopkins text. The file *readme.txt* includes a description of the data.

On pages 412–413, Glass and Hopkins perform several ANOVAs looking for differences across `RACE` on 3 different dependent variables. Reproduce the ANOVA source tables for all 3 analyses. Remember to transform `RACE` to `factor` format, or you will not get the correct results! Also check the variable names to make sure you enter them correctly.

Your results should look like this:

```
             Df  Sum Sq Mean Sq F value Pr(>F)
RACE          3   4.074  1.3580  3.0538  0.028 *
Residuals   596 265.042  0.4447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
             Df Sum Sq Mean Sq F value    Pr(>F)    
RACE          3   6705 2235.04  28.716 < 2.2e-16 ***
Residuals   596  46389   77.83                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)
RACE          3   0.995 0.33161  0.6651 0.5737
Residuals   596 297.156 0.49858               
```