Lab Exercise 7 - Factorial ANOVA

Read the ENTIRE LAB before beginning. TYPE up your results and turn them in on a separate sheet. It does not have to be a formal write-up. Just tell me what you found. Copying and pasting from R is okay ONLY if you have set the font to courier or courier new. Make sure I know what I'm looking at. Due Thursday.

I have data from the entire population of colleges and universities in the United States (source: U.S. News and World Report, 1995). The colleges are classified by two variables:

Region: in what region of the country (according to official U.S. Federal Census Regions) is the college located?
- levels of this variable:
  - S - South (includes Washington D.C.)
  - M - Midwest
  - W - West (includes Alaska and Hawaii)
  - N - Northeast

Public: is this a publicly supported university (i.e., a "state school")?
- levels of this variable:
  - N - no
  - Y - yes

When these variables are crossed into a factorial design, they create eight groups. You are going to take a random sample of 20 colleges from each of these eight groups. Here's how you do it. (Do EXACTLY this!)

```r
> rm(list=ls())
> source("http://ww2.coastal.edu/kingw/psyc480/functions/samplecolleges.R")
> ls()
[1] "sample.colleges"
```

The source() command will fetch a function from the website that will do the sampling for you. The ls() command just confirms that the function has been added to your workspace. Run it as follows. (Do NOT run this function more than once! Every time you run it, you will get a different sample. So before you run this function, make sure you have time to complete the lab. If you come back after a break and run it again, you will get a different sample, and the end of your lab will not match the beginning of your lab.)

```r
> sample.colleges()
FICE Expend Region Public
Min. :1009 Min. : 3379 M:40 N:80
1st Qu.:1742 1st Qu.: 5954 N:40 Y:80
Median :2466 Median : 7364 S:40
3rd Qu.:3443 3rd Qu.: 9512 W:40
Max. :29099 Max. :36704
```

A data frame called CO has been added to your workspace and summarized.

```r
> ls()
[1] "CO" "sample.colleges"
```

The summary I got is different from the summary you will get. It's a random sample, after all. Very carefully running the following command will clean things up a bit for you and make your graphs easier to interpret.

```r
> CO$Region = reorder(CO$Region, CO$Expend, mean)
```
If that seemed to work okay (no error message), remove the sample colleges function from your workspace.

> rm(sample.colleges)

Your workspace should now contain ONLY the CO data frame. You are now ready to begin.

The variable called FICE is a federal identification number for each of the colleges. (FYI, CCU is 3451, which may or may not be in your dataset.) You can ignore FICE. The variable called Expend is the "instructional expenditure per student." I.e., how much money does the college spend per year on your education. This is your dependent variable. Region and Public are explained above and are the IVs. We are interested in how Expend is related to Public and Region. The analysis you will do is:

```
Expend ~ Public * Region
```

Go to it! If you follow the example I did in class, you cannot go wrong. I will answer legitimate questions if you have them, but I will not answer questions that you should know the answer to but don't because you were not in class (or were in class but were playing with your phone instead of paying attention).

Work on your own. ANY EVIDENCE of collaboration between two or more students will result in your getting a zero on this lab.

Things that you will want to include in your report (may not be a complete list--use your best judgment):

1. the summary of the data frame (that printed out when you first got it)
2. the name of the design
3. whether or not the design is balanced
4. if balanced, how many subjects per group
5. the cell means (in a properly aligned table, please)
6. the cell standard deviations (at least 3 decimal places)
7. an evaluation of the homogeneity of variance assumption (is it serious?)
8. the marginal means (both margins) with a statement about main effects
9. an appropriate graph
10. a guess based on the graph as to what effects exist (and why you think so)
11. the ANOVA summary table (properly aligned) with interpretation
12. effect sizes for all effects (significant or not)
13. a post hoc test on the Region variable with interpretation
14. a very brief statement as to why no post hoc test is required on Public
15. a brief summary statement of what you found