STAT502 Lab #1

1. In this part, we will learn some of the basics of the statistical software R. We will learn how to read data in from a file, how to handle variables, how to plot and boxplot data, and how to save these plots.

   (a) Create a directory within your PASS space (X drive) entitled "STAT502". You may also wish to create subdirectories for separate labs, homework, etc.

   (b) Launch R, and change the working directory to STAT502.

   (c) Download "CH16PR07.txt" from Canvas into your STAT502 folder.

   (d) In R, follow File → New script to create an R file to keep track of the commands you'll use for this lab. Include the following lines of code. To run them, simply copy and paste into the R terminal window. You can switch among active windows in R with `ctrl+tab`.

```
p16.7 <- read.table("CH16PR07.txt")
p16.7
p16.7$V1
p16.7$V2
names(p16.7) <- c("improvement","rd","observation")
p16.7
p16.7$improvement
p16.7$imp  #don't need to specify the whole name
p16.7$rd
plot(p16.7$imp)
plot(p16.7$rd,p16.7$imp)
boxplot(improvement~rd,data=p16.7)
p16.7$rd
p16.7$rd <- factor(p16.7$rd,labels=c("low","moderate","high"))
p16.7$rd
boxplot(improvement~rd,data=p16.7)
mean(p16.7$improvement)
sd(p16.7$improvement)
tapply(p16.7$improvement,p16.7$rd,mean)
tapply(p16.7$improvement,p16.7$rd,sd)
p16.7[2,3]; p16.7[,1]; p16.7[2,]; p16.7[,]
```

2. Download the "gpa.dat" data set from Canvas to work through the example we did in class. Use the following command to get the data into R.

```
gpa = read.table("gpa.dat",header=T)
```

(a) Construct side-by-side boxplots for the two groups. Does it appear that there is a significant difference?

(b) For each group, find the sample mean, variance, and sample size. Also, find the pooled sample variance.

(c) Compute the test statistic for $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(d) Use the R functions `qt()` and `pt()` to find the critical value and p-value, respectively, for making a decision about H_0 (assuming $\alpha = .05$). Use the command `?qt` to see documentation for how these functions are used.

(e) Construct a 95% confidence interval for $\mu_1 - \mu_2$.

(f) Now use the R function `t.test()` to carry out the above test. If you use the options correctly (check the documentation), you should get the same answer (it also provides confidence limits).