Fall 2015 - STAT 518 – Project -- Part II
Real Data Analysis: Parametric vs. Nonparametric

The goal of this assignment is to analyze some real data set using both a nonparametric approach and a parametric approach. The data set and research question should be of the type that could be analyzed using the methods we have studied in STAT 518. You should already have gotten data approved by me from your Part 1.

For this final part, you should describe in a typed report of around 4-6 pages:
(1) the individuals in your sample
(2) the variable or variables measured on the individuals
(3) the research question you will answer, including the null and research (alternative) hypotheses
(4) statistical analyses of the data using two approaches for answering the research question (one should be nonparametric and one should be a classical parametric approach)
(5) your specific conclusions about the research question based on your data analysis. Be sure to interpret you results in the context of your real data set.
(6) A plot of two power functions (one for the nonparametric test and one for the parametric test) giving the probabilities of rejecting H_0 both when the null is true and when the alternative is true (and for various values in the "alternative region"). The power calculations are based on simulated data that resemble your real data in some way (see below).
(7) Your conclusions about which approach (nonparametric or parametric) seems preferable, based on the power curves, and why. Do both tests appear to be unbiased?

This part is due by December 9, 2015 (final exam day), but you can turn it in early.

For the final project report, you will not only analyze your data set using both approaches and write up your conclusions, but also perform a simulation study in R to approximate the power of each procedure on simulated data that has similarities to your real data set. This will allow you to make conclusions about which approach is preferable to analyze your data.

Grading:
The project will be graded out of 30 points, of which this final part is worth 20 points. As an encouragement for working in groups, you will get 2 bonus points if you work in a group of two people. When working in groups, each member must contribute significantly to the project.

Power plots:
I have provided some code to calculate power on simulated data for both parametric and nonparametric tests. If the tests you plan to use do not match the examples I have given, you should see me in person. To create a power curve, you can run the code for multiple values of the number (either "mu.d" or "k" in the examples I gave) that specifies how far into the alternative region the truth is. Save the values you used of this number (either "mu.d" or "k") and also save the corresponding powers for both the parametric and nonparametric tests. Then you can plot both power curves against either "mu.d" or "k" values on the same plot for comparison purposes. So you might have something like:

```
my.mu.ds <- c(0, …, …, …)   # fill in with the numbers used…
power.param <- c(0.0501, …, …, …) # fill in with the powers obtained…
power.nonparam <- c(0.0478, …, …, …) # fill in with the powers obtained…
plot(my.mu.ds, power.param, type="l", xlab="mu.d", ylab="power") # gives solid lines
lines(my.mu.ds, power.nonparam, lty=2) # gives dashed lines
abline(h=0.05, lty=3) # shows horizontal line at alpha
```