STAT 672, Midterm Exam Part II

Due 4:00PM, Wednesday March 22

This assignment is a individual take-home midterm exam. You **may not** collaborate with *any* other person (whether in the class or not). You **may** use any reading material (class notes, books, etc.) you wish. Professor Bailey will answer questions. Please include R commands and output you used to answer the questions.

2 Problems. 25 points total. Show all work.

(10pts) 1. Let X have a Cauchy distribution and let Y have a standard normal distribution. I have simulated a random sample from each of these distributions to use as two datasets. I have used the R function `rcauchy(100)` and `rnorm(100)`, in case you would like more information. You can find these two data sets in a file off the course webpage and you can use the read.table command:

```
testdat <- read.table("http://www.rohan.sdsu.edu/~babailey/stat672/testdat.txt",
header=T)
```

Note: The names of the variables in the file are x and y.

(a) First consider the random sample from the Cauchy distribution ($X's$). Use the Wilcoxon Sign Rank test to test if the population median is equal to zero against the two-sided alternative. What do you conclude at the $\alpha = 0.10$ significance level? In addition, estimate θ and sort through the Walsh averages to find an approximate 90% CI for θ. What is the achieved confidence level? Are the assumptions of the Wilcoxon Sign Rank test satisfied? Explain.

(b) You should have two random samples of size $n = m = 100$ from two distributions ($X's$ and $Y's$). Use the Wilcoxon Rank Sum procedure to test $H_0 : \Delta = 0$ against the two-sided alternative. What do you conclude at the $\alpha = 0.10$ significance level? Are the assumptions of the Wilcoxon Rank Sum test satisfied? Explain.

(15pts) 2. Let $X_1, X_2, \ldots, X_9$ be random sample of size $n = 9$ from a continuous distribution. Suppose we question whether they are observations of a random sample, in fact we suspect a trend. Let $R_i = \text{rank}(X_i)$ and take $a_1 = a_2 = a_3 = 1, a_4 = a_5 = a_6 = 2, a_7 = a_8 = a_9 = 3$. A statistic that could be used to test the alternative (trend) hypothesis is

$$L = \sum_{i=1}^{9} a_i R_i.$$

(a) Under the assumption (H_0) that the n random variables are actually observations from a random sample from a continuous distribution, find $\mu_L = E(L)$ and $\sigma_L^2 = Var(L)$.

(b) Consider the following sequence of observations,

```
> x
 [1] 1.8 1.0 1.5 1.4 1.9 1.6 1.2 1.7 1.3
```

The rejection region is of the form $L \geq c$. Use the normal approximation to calculate the approximate p-value for the test. What is your conclusion at the $\alpha = 0.05$ significance level?