Math 159
Nonparametric Analysis
Spring 2010
Jo Hardin
HW6: Section 5.1

HOMEWORK THOUGHTS: A large part of this course is learning how to interpret and communicate results. That is, an isolated p-value will not ever be a complete answer to a question. As stated below, please always explain your answers in a sentence or two (unless, of course, the problem is truly just computational).

1. Section 5.1: 3, 4, 6b
2. Consider the $(1-\alpha)$ 100% CI derived in section 5.1. Show that when $\alpha = 2/\binom{N}{n}$ (really, your α will be slightly larger, so $\alpha = 2/\binom{N}{n} + \epsilon$, where $\epsilon > 0$ is some tiny amount),

$$\begin{aligned} L &= X^{(1)} - Y^{(m)} \\ U &= X^{(n)} - Y^{(1)} \end{aligned}$$

   where $X^{(1)} \leq \ldots \leq X^{(n)}$ are the ordered X's, and $Y^{(1)} \leq \ldots \leq Y^{(m)}$ are the ordered Y's.
3. Section 5.4: Exercises: 1 (you can use Excel, but don't use R), 2 (use R), 5
4. Section 5.4: Problems: 2, 4
5. Find the dataset softdrink.txt on my website. The dataset represents 25 observations on delivery time (min), number of cases, and distance walked (feet) by a soft drink company's service routes in the distribution system.

   Analyze the correlation among these three variables. That is, compute the Pearson, Spearman, and Kendall correlations. Give significance levels (p-values) for all three, and a CI for the Pearson correlation. Then discuss any assumptions that may or may not be valid. (Recall that the distribution of the Pearson correlation is based on the assumption that the variables have (bivariate) normal distributions and that the Pearson correlation is not robust to outliers.)

   - To read the file into R, first put the dataset in the same folder (possibly called Statistics on your user space) as R.
     `> softdrink<-read.table("softdrink.txt",header=T,sep="\t")`
   - To measure correlation use the `cor` function with the option `method="p"`, `method="s"`, or `method="k"` for the correlation of interest. Note that R understands an entire matrix (and will find all the pairwise correlations simultaneously.)
   - To test correlation, use `cor.test` and the same method commands as above for the distinct measures.
   - To check for assumptions, you might try histograms (`hist`), pairwise scatter plots (`pairs, plot`), and removing outlying points.