## Assignment 6 - CSC/DSC 262/462 - Fall 2017 - Due December 13

**Q1:** A weight loss program is given to $n = 7$ participants of age 11 years. The weight in pounds for each subject before and after the program are given in the following table. Suppose $\tilde{\mu}_D$ is the population median of the paired differences $D = Y - X$. Perform a lower tailed signed rank test using hypotheses $H_o : \tilde{\mu}_D = 0$ against $H_a : \tilde{\mu}_D < 0$. Use significance level $\alpha = 0.05$. Use both the exact signed rank distribution and the normal approximation. Verify your answer using the `wilcox.test()` function. Do not use any continuity correction.

| | Before Program ($X$) | After Program ($Y$) | Difference ($D = Y - X$) |
|---|---|---|---|
| 1 | 91.4 | 85.6 | -5.8 |
| 2 | 101.1 | 94.6 | -6.5 |
| 3 | 97.7 | 101.8 | +4.1 |
| 4 | 95.5 | 87.8 | -7.7 |
| 5 | 100.7 | 96.6 | -4.1 |
| 6 | 102.6 | 105.5 | +2.9 |
| 7 | 84.0 | 89.1 | +5.1 |

**Q2:** We are given two independent samples of sample sizes $n_1 = 7$, $n_2 = 10$. The data is summarized in the table below. Suppose $\tilde{\mu_i}$ is the population median of sample $i$. Perform a two-sided rank sum test using hypotheses $H_o : \tilde{\mu_1} - \tilde{\mu_2} = 0$ against $H_a : \tilde{\mu_1} - \tilde{\mu_2} \neq 0$. Use significance level $\alpha = 0.05$. Use the normal approximation method only. Verify your answer using the `wilcox.test()` function. Do not use the continuity correction in either case. What method does `wilcox.test()` use when there are ties?

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\tilde{X}_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample 1 | 31.7 | 20.9 | 23.2 | 30.2 | 34.4 | 31.0 | 26.9 | | | | 30.2 |
| Sample 2 | 35.0 | 40.8 | 39.8 | 30.2 | 40.9 | 32.7 | 38.9 | 35.3 | 35.2 | 38.9 | 37.1 |

**Q3:** A random sample of categorical data is collected. The sample is of size $n = 124$. The observed counts for $k = 4$ categories are given in the following table. Use a $\chi^2$ test for the null hypothesis

$$H_o : p = (1/15, 2/15, 4/15, 8/15).$$

Use significance level $\alpha = 0.05$. Do not use Yate's correction. Verify your answer using the `R` function `chisq.test()`

| | 1 | 2 | 3 | 4 | Totals |
|---|---|---|---|---|---|
| Observed counts $O_i$ | 8 | 19 | 31 | 66 | 124 |
| Hypothetical frequencies $p_i$ | 1/15 | 2/15 | 4/15 | 8/15 | 1.00 |
| Observed frequencies $\hat{p}_i$ | - | - | - | - | 1.00 |

**Q4:** It has been reported that left-handedness is more common in males. Suppose the following table summarizes data from $n = 6500$ individuals categorized as male/female and left/right handedness. Hypothetical population frequencies of cell $i, j$ are given by $p_{i,j}$. The population frequencies for row $i$ and column $j$ categories are given by $r_i$ and $c_j$, respectively. Use a $\chi^2$ test for the null hypothesis of row and column independence: $H_o : p_{i,j} = r_i c_j$ for all $i, j$. Use significance level $\alpha = 0.05$. Use Yate's correction. Verify your answer using the `R` function `chisq.test()`. How is your conclusion related to the conjecture that left-handedness is more common in males?

**Q5:** **[For graduate students]** We wish to investigate the accuracy of the $t$-distribution approximation for the transformed correlation of Section 21.1 when applied to the Spearman rank correlation coefficient.

| | Right Handed | Left Handed | Totals |
|---|---|---|---|
| Male | 2597 | 425 | 3022 |
| Female | 3128 | 350 | 3478 |
| Totals | 5725 | 775 | 6500 |

Table 1: Observed counts $O_{i,j}$

(a) The `R` function `sample(n)` outputs a random permutation of the numbers $1, 2, \ldots, n$. How can this be used to simulate the Spearman rank correlation coefficient when the true correlation is $\rho = 0$?

(b) Simulate a random sample (of size 10,000) of Spearman rank correlation coefficients for $n = 25$ under the null hypothesis $H_o : \rho = 0$.

(c) Calculate the $t$-distribution transformation for each simulated value (Equation (21.1) of Section 21.1).

(d) Plot a histogram of the transformed sample. Superimpose a $t$-density on the same plot, using the approriate degrees of freedom. Is the $t$-distribution an accurate approximation? [Make sure you use the `probability = T` option when you plot the histogram].