1. (5 points) Tony collected daily sleeping time for m = 4 coffee drinkers and n = 4 tea drinkers. He reported that, according to the Wilcox rank-sum test on the data set, the average daily sleeping time of coffee drinkers differs from the average daily sleeping time of tea drinkers at α = 0.01 level. His statistics professor tells Tony that this conclusion is impossible mathematically.

(a) Can the Wilcox rank-sum test results in a p-value less than 0.01 on such a data set? If no, why not? If yes, can you make up a such data set with m = 4 and n = 4?

(b) If Tony used the permutation test for sample means difference instead of the Wilcox rank-sum test, would he be able to get a p-value less than 0.01 on a data set of size m = 4 and n = 4?

Solution: Given, Tony collected daily sleeping time for m = 4 coffee drinkers and n = 4 tea drinkers.

(a) The minimal sample size needed for the Wilcox rank-sum test to produce a p-value less than 0.01 is (m + n) = 13, so a data set with m = 4 and n = 4 cannot utilize the Wilcox rank-sum test to produce a p-value of less than 0.01.
The Wilcox rank-sum test is based on comparing the sum of the ranks of the smaller of the two groups to the expected value of this sum under the null hypothesis, which follows a normal distribution. With m = 4 and n = 4, the smallest value possible p-value that can be obtained from the test is 0.0625, which is greater than 0.01.
The normal approximation only holds true for samples with greater sizes. Thus, it is impossible for the Wilcox rank-sum test on the given data to produce a p-value of less than 0.01.

(b) On a set of data with size m = 4 and n = 4, it is possible for the permutation test for sample means difference to have a p-value less than 0.01. In actuality, the following data set may be created:
Coffee drinkers: 6, 7, 8, 9
Tea drinkers: 1, 2, 3, 10.
The mean sleeping time for the coffee drinkers is, (6 + 7 + 8 + 9)/4 = 7.5 hours.
The mean sleeping time for the tea drinkers is, (1 + 2 + 3 + 10) = 4 hours.
The difference in means is 7.5 - 4 = 3.5 hours.
All possible permutations of the data are equally likely under the null hypothesis that there is no difference in the mean sleeping time between coffee and tea drinkers.

There are 70 possible permutations to combine the data, but only one (6, 7, 8, 9, 1, 2, 3, 10) yields a different mean that is bigger than or equal to 3.5 hours.

Thus, the permutation test's p-value is 1/70, which is less than 0.01 for significance.

Because the rankings given to the data do not accurately reflect the size of the differences between the observations, the Wilcox rank sum-test would not be able to identify this discrepancy in means.

Therefore, Tony's statistics professor is right that a Wilcox rank-sum test with a size of m = 4 and n = 4 cannot have a p-value less than 0.01.

Using such data, the permutation test for sample means difference could yet produce a p-value less than 0.01.

2. (10 points each)

Exercise 13.7.7. We are interested in examining the effects of the transition from fetal to postnatal circulation among premature infants. For each of 14 healthy newborns, respiratory rate is measured at two different times: once when the child is less than 15 days old, and again when he or she is more than 25 days old [227]. The data are presented below.

SUBJECT	RESPIRATORY RATE (BREATHS/MINUTE)	
	Time 1	Time 2
1	62	46
2	35	42
3	38	40
4	80	42
5	48	36
6	48	46
7	68	45
8	26	40
9	48	42
10	27	40
11	43	46
12	67	31
13	52	44
14	88	48

(a) Using the sign test, evaluate the null hypothesis that the median difference in respiratory rates for the two times is equal to 0. Conduct the test at the 0.05 level of significance. What do you conclude?

(b) Evaluate the same hypothesis using the Wilcoxon signed-rank test. What do you conclude?

Solutions: (a) Null Hypothesis and Alternative Hypothesis are as follows:

$H_0: Median\ Difference\ =\ 0$

$H_a: Median\ Difference\ \neq\ 0$

$\alpha\ =\ 0.05$

TIME 1	TIME 2	DIFF (TIME 1 – TIME 2)	SIGN
62	46	16	+
35	42	-7	-
38	40	-2	-
80	42	38	+
48	36	12	+
48	46	2	+
68	45	23	+
26	40	-14	-
48	42	6	+
27	40	-13	-
43	46	-3	-
67	31	36	+
52	44	8	+
88	48	40	+

Positive Sign Count = 9

Negative Sign Count = 5

Total Count = 14

Z-score Calculation:

$$z = \frac{(X - pn)}{\sqrt{npq}}$$

$$z = \frac{(9 - 7)}{\sqrt{3.5}} = 1.06904$$

The z – score is 1.06904. The p-value is 0.28505. The result is not significant because $p > 0.05$, and therefore we fail to reject the null hypothesis.

(b)

Subject	***Time 1***	***Time 2***	***Abs Diff***	***Rank***	***Sign***
3	38	40	2	1.5	-1
6	48	46	2	1.5	+1
11	43	46	3	3	-1
9	48	42	6	4	+1
2	35	42	7	5	-1
13	52	44	8	6	+1
5	48	36	12	7	+1
10	27	40	13	8	-1
8	26	40	14	9	-1
1	62	46	16	10	+1
7	68	45	23	11	+1
12	67	31	36	12	+1
4	80	42	38	13	+1
14	88	48	40	14	+1

The sum of positive ranks is:

$$W+ = 1.5 + 4 + 6 + 7 + 10 + 11 + 12 + 13 + 14 = 78.5$$

The sum of negative ranks is:

$$W- = 1.5 + 3 + 5 + 8 + 9 = 26.5$$

Hence, test statistics $T = min\{W+, W-\} = min\{78.5, 26.5\} = 26.5$

Null Hypothesis and Alternative Hypothesis are as follows:

H_0: $Median\ Difference = 0$

H_a: $Median\ Difference \neq 0$

The critical value for the significance level α=0.05 provided, and the type of tail specified is $T * = 21$, and the null hypothesis is rejected if $T \leq 21$.

Since in this case $T = 26.5 > 21$, there is not enough evidence to reject the null hypothesis at the $\alpha = 0.05$ significance level.

Exercise 13.7.11. A study was conducted to evaluate the effectiveness of a work site health promotion program in reducing the prevalence of cigarette smoking. Thirty-two work sites were randomly assigned to either implement the health program or to make no changes for a period of two years. The promotion program consisted of health education classes combined with a payroll-based incentive system. The data collected during the study are saved in the dataset *work_program* [229]. For each work site, smoking prevalence at the start of the study is saved under the variable name baseline, and smoking prevalence at the end of the two-year period under the name *followup*. The variable group contains the value 1 for the work sites that implemented the health program and 2 for the sites that did not.

(a) For the work sites that implemented the health promotion program, test the null hypothesis that the median difference in smoking prevalence over the two-year period is equal to 0 at the 0.05 level of significance. What do you conclude?

(b) Test the same null hypothesis for the sites that did not make any changes. What do you conclude?

(c) Evaluate the null hypothesis that the median difference in smoking prevalence over the two-year period for work sites that implemented the health program is equal to the median difference for sites that did not implement the health program. Again, conduct the test at the 0.05 level of significance. What do you conclude?

(d) Do you believe that the health promotion program was effective in reducing the prevalence of smoking? Explain.

Solution:

```
> library(haven)

> work <- read_dta("C:/Users/abhil/OneDrive/Desktop/MSAM-Northeastern/MATH7343/Homeworks/6/work_program.dta")

> attach(work)

The following objects are masked from work (pos = 3):

    baseline, followup, group

The following objects are masked from work (pos = 4):

    baseline, followup, group

The following objects are masked from work (pos = 5):

    baseline, followup, group
```

(a)

```
> # ------------- (a) --------------------
> x1 = baseline[group == 1]
> y1 = followup[group == 1]
> wilcox.test(x1, y1, paired = T)

        Wilcoxon signed rank test with continuity correction

data:  x1 and y1
V = 115, p-value = 0.001966
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x1, y1, paired = T) :
  cannot compute exact p-value with zeroes
```

We reject the null hypothesis because the *p-value* is so small and conclude that the median difference in smoking prevalence over the course of the promotion program's implementation and work sites' two-year timeframe is substantially different from 0.

(b)

```
> # ------------- (b) --------------------
> x2 = baseline[group == 2]
> y2 = followup[group == 2]
> wilcox.test(x2,y2,paired = T)

        Wilcoxon signed rank exact test

data:  x2 and y2
V = 86, p-value = 0.3755
alternative hypothesis: true location shift is not equal to 0
```

Since p-value > 0.05, we accept the null hypothesis and conclude that the median difference in smoking prevalence over the course of the two-year period for the workplaces not using the promotion program is not substantially different from 0.

(c)

```
> # ------------- (c) --------------------
> diff1 = x1 - y1 #Group1
> diff2 = x2 - y2 #Group2
> wilcox.test(diff1, diff2, paired = F)

        Wilcoxon rank sum exact test

data:  diff1 and diff2
W = 177, p-value = 0.06708
alternative hypothesis: true location shift is not equal to 0
```

Since p-value > 0.05, we accept the null hypothesis and conclude that there was little benefit from the health program since median difference in both the cases is not substantially different from 0.

(d) Since p-value > 0.05, we accept the null hypothesis and conclude that there was little benefit from the health program because the median difference in smoking prevalence over the two-year period for workplaces that implemented the program was not significantly different from the median difference for workplaces that did not.

Exercise 13.7.12. A study was conducted to investigate whether females who do not have health insurance coverage are less likely to be screened for breast cancer than those who do, and whether their disease is more advanced at the time of diagnosis [230]. The medical records for a sample of individuals who were privately insured and for a sample who were uninsured were examined. The stage of breast cancer at diagnosis was assigned a number between 1 and 5 where 1 denotes the least advanced disease and 5 the most advanced. The relevant observations are saved in the dataset called *insurance*; the stage of disease is saved under the variable name *stage*, and an indicator of group status - which takes the value 1 for females who were uninsured and 0 for those who were privately insured - under the name group.

(a) Could the two-sample t test be used to analyze these data? Why or why not?

(b) Test the null hypothesis that the median stage of breast cancer for females with private insurance is identical to the median stage of cancer for those who are not insured.

(c) Do these data suggest that uninsured females have more advanced disease at the time of diagnosis of breast cancer than those who are insured? Explain.

Solution: (a) The sample doesn't appear to be normally distributed, hence t-test is not the appropriate test.

(b)

```
> insured_women <- subset(insure, insure$group == 0)
> uninsured_women <- subset(insure, insure$group == 1)
> wilcox.test(insured_women$stage, uninsured_women$stage)

        Wilcoxon rank sum test with continuity correction

data:  insured_women$stage and uninsured_women$stage
W = 28758, p-value = 1.496e-05
alternative hypothesis: true location shift is not equal to 0
```

$$p-value = 1.496e-05 < \alpha = 0.05$$

Hence, we reject the null hypothesis and conclude that the median of insured women is not equal to median of uninsured women.

(c) From the data, it is clear that median stage of uninsured data is more than the median stage of insured data, and could be concluded that uninsured females have more advanced disease at the time of diagnosis of breast cancer than those who are insured.

3. (15 points) Do exercise 13.7.13 on page 21. Also, program a permutation-test version of the test used in this problem. Compare the result using your permutation-test code versus your answer in (d).

Exercise 13.7.13. The nursing home occupancy rates for each state in the United States in 2015 are saved in the dataset *nursing_home* [15]. For each state, the variable occupancy is the percentage of beds occupied (number of nursing home residents per 100 nursing home beds), and region contains the region of the United States for that state, defined as Northeast, Southeast, Midwest, Southwest, and West.

(a) Construct box plots of the nursing home occupancy rates by region of the United States.

(b) What test would you use to evaluate the null hypothesis that median nursing home occupancy rates are the same in each region of the country? Explain your choice.

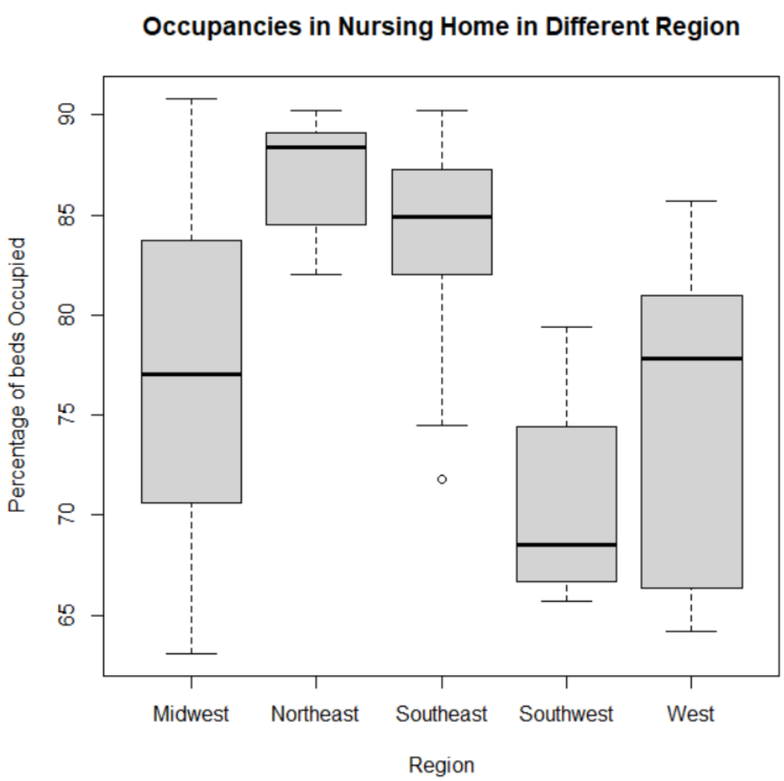
(c) What is the probability distribution of the test statistic for this technique?

(d) Carry out the test. What is the p-value? Do you reject or fail to reject the null hypothesis at the 0.05 level of significance?

(e) What do you conclude?

(f) If you wish to characterize differences in nursing home occupancy rates in different regions of the United States, is there another step that needs to be taken in this analysis? Explain.

Solution: (a)



(b) To evaluate the null hypothesis that median nursing home occupancy rates are the same in each region of the country, we can use the Kruskal-Wallis test. This test is a nonparametric alternative to the one-way ANOVA test, and it does not assume that the data are normally distributed. The Kruskal-Wallis test assesses whether the medians of the groups are statistically different.

(c) The Kruskal-Wallis test statistic follows a chi-squared distribution with k - 1 degrees of freedom, where k is the number of groups (regions) being compared.

(d) RStudio Result of Kruskal-Wallis Test:

```
> kruskal.test(occupancy ~ region, data = nursing)

        Kruskal-Wallis rank sum test

data:  occupancy by region
Kruskal-Wallis chi-squared = 18.43, df = 4, p-value = 0.001017
```

The p-value is 0.001017, which less than 0.05, which means we reject the null hypothesis that the median nursing home occupancy rates are the same in each region of the country.

(e) We can conclude that there is significant evidence that the median nursing home occupancy rates differ among regions of the United States.

(f) To characterize the differences in nursing home occupancy rates in different regions of the United States, we can perform post-hoc tests to determine which pairs of regions have significantly different median occupancy rates. A common post-hoc test for the Kruskal-Wallis test is the Dunn's test, which compares all possible pairs of groups using a Bonferroni correction for multiple comparisons. Alternatively, we could use the Wilcoxon rank-sum test (also known as the Mann-Whitney U test) to compare pairs of regions.