Statistical Method

Homework 4

For each question, please select suitable testing and give the reason (could be multiple suitable testing). Analyze the data and give the conclusion regarding to the interest of each question.

1. We collect the scores of the MATH exam from two classes, Class A and Class B. There are 25 students in Class A and 21 students in Class B. The teacher wants to know if the students performed equally on the exam. The dataset is shown in the sheet Question 1.

   The purpose is to compare if the centrals or distributions from A and B are the same. The students are independent in and between both Class A and Class B. The null hypothesis is

   $$H_0 : \mu_A = \mu_B$$

   Due to the sample sizes are 25 and 21, respectively, do the examination of the data first in Figure 1, and they are not like the normal distributed shapes. Thus, I will use the non-parametric way to test if the scores of two classes are the same. Figure 2 shows the result of the Mann-Whitney U Test, and the p-value of the test is 0.015 less than 0.05. Hence, we reject the null hypothesis, which means the data has the evidence to show the scores between Class A and Class B are significantly different.

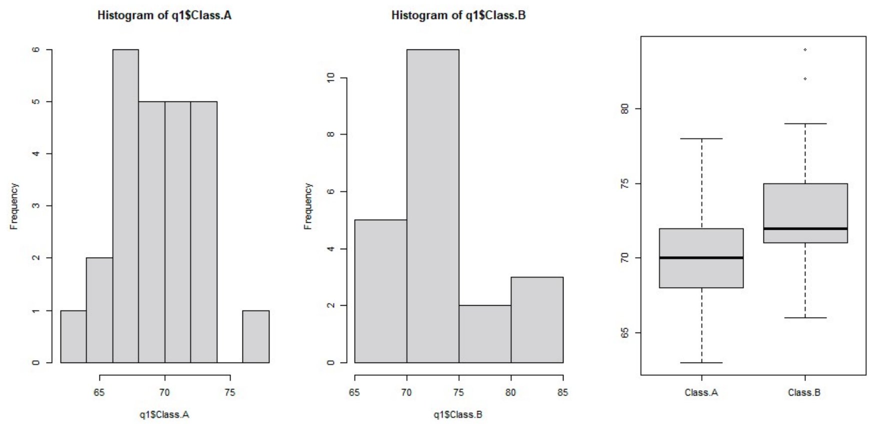

Figure 1: EDA for Question 1

```
> wilcox.test(q1$Class.A, q1$Class.B)

        Wilcoxon rank sum test with continuity correction

data:  q1$Class.A and q1$Class.B
W = 152, p-value = 0.01484
alternative hypothesis: true location shift is not equal to 0
```

Figure 2: Mann-Whitney U test for Question 1

2. Twelve farmers are selected randomly in an experiment with a plant nursery. Each farmer is asked to select four fair and identical areas in the yard and to plant four different types of grasses. One type in each area of the yard. After the grasses grow, the farmer will give the score of each grass. The experiment is conducted to know if the four types of grasses are popular equally. The dataset is shown in the sheet Question 2.

   The purpose is to know if the four types of grasses are popular equally. There are 12 farmers providing their opinion among the four types of grasses. That is, there are four measurements made by one farmer, and it is called repeated measures. We should assume that the measurements among farmers are independent, but not independent within scores of four types of grasses for each farmer. In addition, the scores are from 1 to 4, which could violate the assumption of the normal distribution in Figure 3. Hence, I choose to use the non-parametric method, Friedman test, to test if the four types of grasses are popular equally. From Figure 4, I set the judge as the block effect and the type of grasses is the group effect (main concerned). The result shows that the p-value is 0.044 less than 0.05, and it reject the null hypothesis, $H_0 : \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{T4}$. It shows the four types of grasses are not popular equally.

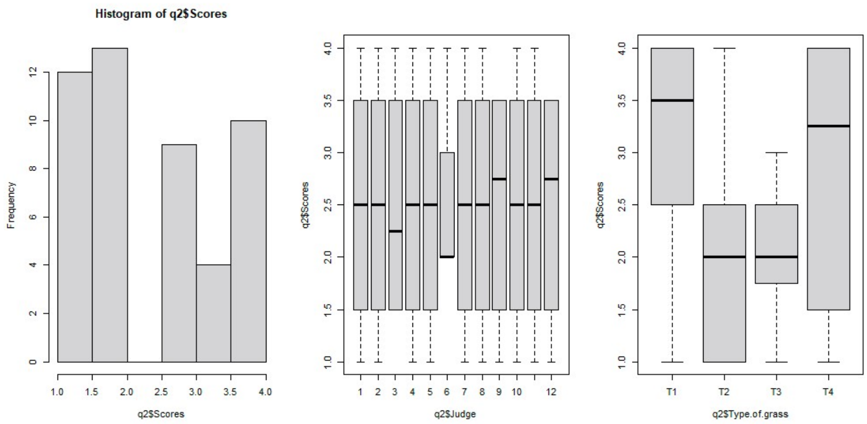

Figure 3: Mann-Whitney U Test for Question 1

Alternatively, if we ignore the effect from the judges, we can use Kruskal-Wallis test. The result shown in Figure 5 provide the evidence to reject the null hypothesis (p-value = 0.015). It shows the four types of grasses are not popular equally.

```
> q2 <- read.csv("HW_Q2.csv")
> friedman.test(q2$Scores, groups = q2$Type.of.grass, blocks = q2$Judge)

        Friedman rank sum test

data:  q2$Scores, q2$Type.of.grass and q2$Judge
Friedman chi-squared = 8.0973, df = 3, p-value = 0.04404
```

Figure 4: Friedman test for Question 2

```
> kruskal.test(Scores ~ Type.of.grass, data = q2)

        Kruskal-Wallis rank sum test

data:  Scores by Type.of.grass
Kruskal-Wallis chi-squared = 10.446, df = 3, p-value = 0.01513
```

Figure 5: Kruskal-Wallis test for Question 2

3. 50 people were surveyed regarding their opinion about candidates for Mayor. 15 people were in Candidate A and 35 people were in Candidate B. After they listened to a debate by the two candidates and the survey done by the 50 people was repeated. Then, 17 voted in Candidate A and 33 in Candidate B. Did the debate affect people's opinions? The dataset is shown in the sheet Question 3.

   The people are the same before and after the debate, and they all give the opinion of supporting. The purpose is to see if the debase could change their mind. The McNemar's test is to see if the probabilities of changing mind before and after the debate are the same. Figure 6 shows the p-value is 0.723 larger than 0.05. Hence, the data has no evidence to reject the null hypothesis that the probabilities of changing mind are the same. That is, there is no evidence to show the debate affected people's opinions.

```
> q3 <- read.csv("HW_Q3.csv")
> table.q3 <- table(q3[,2:3])
> table.q3
                 After.the.debate
Before.the.debate  A  B
                A 12  3
                B  5 30
> mcnemar.test(table.q3)

        McNemar's Chi-squared test with continuity correction

data:  table.q3
McNemar's chi-squared = 0.125, df = 1, p-value = 0.7237
```

Figure 6: McNemar's test for Question 3