Statistical methods
Homework 4

RE6124019 Matthew

Due data: 23:59, October 23, 2023

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.

```
> #1
> # Welch's t test
> t.test(dat1$`Class A`, dat1$`Class B`, alternative = "two.sided")

        Welch Two Sample t-test

data:  dat1$`Class A` and dat1$`Class B`
t = -2.7216, df = 32.574, p-value = 0.01035
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.1327273 -0.8844155
sample estimates:
mean of x mean of y
 69.92000  73.42857
```

Figure 1: Welch's t test

(a) Hypothesis Setup:

Null Hypothesis (H_0):

The average scores of students in Class A and Class B in the math exam are equal.

$$H_0 : \mu_{\text{Class A}} = \mu_{\text{Class B}}$$

Alternative Hypothesis (H_1):

The average scores of students in Class A and Class B in the math exam are not equal.

$$H_1 : \mu_{\text{Class A}} \neq \mu_{\text{Class B}}$$

(b) Choice of Statistical Test

In this scenario, an independent samples t-test is chosen since there are two independent samples (Class A and Class B), and we are comparing the means of these two groups.

(c) Conducting Independent Samples t-test

In the R code, the `t.test` function was used to perform an independent samples t-test, providing the t-statistic and p-value, as depicted in Figure 1.

(d) Decision Making

p-value = 0.01035:

- Since $0.01035 < 0.05$, we reject the null hypothesis.
- Therefore, there is sufficient evidence to conclude that the average scores in the two classes are statistically different.

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.

```
> #2
> # Friedman test
> friedman.test(y=dat2$Scores, groups=dat2$`Types`, blocks=dat2$Judge)

        Friedman rank sum test

data:  dat2$Scores, dat2$Types and dat2$Judge
Friedman chi-squared = 8.0973, df = 3, p-value = 0.04404
```

Figure 2: Kruskal-Wallis test

(a) Hypothesis Setup:

Null Hypothesis (H_0):

The median scores of the four grass types (T1, T2, T3, T4) are equal.

$$H_0 : \mu_{\text{T1}} = \mu_{\text{T2}} = \mu_{\text{T3}} = \mu_{\text{T4}}$$

Alternative Hypothesis (H_1):

The median scores of at least one of the grass types is different from the rest.

$$H_1 : \mu_i \neq \mu_j, \text{ for at least one pair (i,j) of grass types}$$

(b) Choice of Statistical Test:

The Friedman test is chosen since there are 12 farmers each scoring 4 grass types. This is a repeated measures design with the same subjects exposed to all treatments. The Friedman test is appropriate for comparing medians of k dependent/related samples.

(c) Conducting Friedman Test:

The Friedman test ranks the rows, sums the ranks by column, and compares the mean ranks between columns using a chi-squared statistic.as shown in Figure 2

(d) Decision Making:

p-value = 0.04404:

Since the p-value $= 0.04404 < 0.05$, we reject the null hypothesis.

There is evidence that the median scores for the 4 grass types are not all equal. At least one grass type has a different median score.

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.

```
> #3
> # Contingency table as matrix
> df <- matrix(c(12, 3, 5, 30), nrow = 2, byrow = TRUE,
+              dimnames = list(c('A','B'), c('A','B')))
>
> # McNemar's test
> mcnemar.test(df)

        McNemar's Chi-squared test with continuity correction

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

Figure 3: McNemar's test

1. Hypothesis Setup:

   Null Hypothesis (H_0):

   There is no difference in the proportions of voters for candidates A and B before and after the debate.

$$H_0 : p_{\text{before}} = p_{\text{after}}$$

   Alternative Hypothesis (H_1):

   There is a difference in the proportions of voters for candidates A and B before and after the debate.

$$H_1 : p_{\text{before}} \neq p_{\text{after}}$$

2. Choice of Statistical Test:

   The McNemar's test is chosen since there are two paired nominal samples (before and after the debate) and we want to compare the proportions choosing candidate A or B.

3. Conducting McNemar's test:

   The `mcnemar.test` in R provided a chi-squared statistic of 0.125 and a p-value of 0.7237, as shown in Figure 3:

4. Decision Making:

   p-value = 0.7237

   - Since $0.7237 > 0.05$, we do not reject the null hypothesis.
   - There is not sufficient evidence that the proportions of voters changed between before and after the debate.