

test_s7

Intro

The focus of this practice session will be to perform hypothesis tests for the difference of two or more means. We will look at various test statistics that can be used. We will also conduct hypothesis testing for correlations.

Question 1: Exercise Hypothesis test for the difference of two means

A study is interested to check if the mean exercise hours for female are less than the mean exercise hours for male students. Use data `ExerciseHours` and the two variables `Exercise` and `Sex`.

1.) **Step 1:** Write the null hypothesis and alternative hypothesis in words and in symbols.

a.) Create a boxplot to describe hours of exercise for `female` versus `male`.

```
# your code here
```

b.) Find some favorites statistics of `Exercise` hours for female and male students. You might find the function: `mosaic::favstats` useful. *Note:* you can search online for this function arguments.

```
#your code here
```

c.) Subset the data `ExerciseHours` to two groups: F and M.

```
#your code here
```

2.) **Step 2:** Compute the observed statistic (mean difference of exercise hours for Female and Male).

```
#your code here
```

3.) **Step 3:** Create null hypothesis distribution

a.) Shuffle the two groups of **female** and **Male** into two samples, and find the mean difference of the two shuffled samples.

b.) Create the Null hypothesis Distribution using `do_it()` function.

c.) Plot a **histogram** of the null distribution and show the line of the observed mean difference using the `abline()` function.

```
# your code here
```

4.) **Step 4:** Calculate p-value

```
# your code here
```

Step 5: Make decision/Judgment

```
#your code here
```

Question 5:

We will use data on **Antibodies** (in g/ml) production after receiving a **Vaccine** (Vaccine A, Vaccine B, Vaccine C). A hospital administered three different vaccines to 6 individuals each and measured the antibody presence in their blood after a chosen time period. The data is saved in `patient_vaccine.csv`.

We walk you through testing for the difference between the three groups of vaccines using a different method than in class, it is called the Kruskal-Wills test.

- 1) Create a boxplot to show the three vaccines variation in terms of the antibodies.

```
#your code here
```

- 2.) Write in words the **null hypothesis** and the **alternative hypothesis**.

```
#your code here
```

- 3.) Let prepare your data. Rank your data from all groups together in one column, name it **ranks**. *hint*: you can use function **rank**.

```
#your code here
```

- 4.) Sum the ranks for each group of the **Vaccine**. Reports those sums results.

```
#your code here
```

- 5.) Calculate the test statistic, H of the Kruskal-Wills test given by the formula:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Where :

- N is the total sample size

- k is the number of groups we are comparing.
- $\sum R_i$ is the sum of ranks for group i .
- n_i is the sample size of group i .

```
#your code here
```

6.) Compare the test statistics H to the critical cutoff determined by the critical value **chi-square**. *hint:* from the chi-square table, find the chi-square **critical value** with degrees of freedom $df = k - 1$.

```
#your code here
```

7.) Make Judgement about your hypothesis within the context.

```
#your code here
```

Question 6:

Repeat Question 5 with the randomization method **MAD** and compare your results. What is your reflection.

```
#your code here
```