Drug_test_project_Assignment_Chi

2022-06-28

Introduction:

The data of this project assignment is the "placebo_new_drug" data which is about results from testing new drug. The sample is small with 5 objects. Hence, I will choose non-parametric test to analyze the data.

The Analysis:

Firstly, I import the libraries and the data:

```
library(data.table)
library(BSDA)
```

```
## Loading required package: lattice

##
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':
##
##     Orange
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
setwd("/Users/kienguyen/Documents/DATA SCIENCE/MSDS/03. MSMS 660/08. Week 8/In class")

drugdt <- read.csv(file = 'placebo_new_drug.csv',sep=",", header=T)
```

Next, I calculate the median of the 2 groups:

```
median(drugdt$Placebo)
```

```
## [1] 6
```

```
median(drugdt$New.Drug)
```

```
## [1] 3
```

The mean of "Placebo" variable is greater than the "New.Drug" variable, hence, I will use "greater" alternative for the below tests about the means.

Test the difference between medians of 2 groups:

SIGN test:

```
SIGN.test(x = drugdt$Placebo, y = drugdt$New.Drug , alternative = 'greater')
```

```
##
##  Dependent-samples Sign-Test
##
## data:  drugdt$Placebo and drugdt$New.Drug
## S = 4, p-value = 0.1875
## alternative hypothesis: true median difference is greater than 0
## 95 percent confidence interval:
##  -0.64   Inf
## sample estimates:
## median of x-y
##             2
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.8125   2.00    Inf
## Interpolated CI       0.9500  -0.64    Inf
## Upper Achieved CI     0.9688  -1.00    Inf
```

p-value = 0.1875 >0.05. We fail to reject the null hypothesis and state that the medians are NOT significantly different.

Wilcoxon Signed Rank test

```
wilcox.test(drugdt$Placebo, drugdt$New.Drug, alternative = 'greater')
```

```
## Warning in wilcox.test.default(drugdt$Placebo, drugdt$New.Drug, alternative =
## "greater"): cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  drugdt$Placebo and drugdt$New.Drug
## W = 22, p-value = 0.02928
## alternative hypothesis: true location shift is greater than 0
```

We received the red warning because of the small size sample. So, with the p-value = 0.02928 $<$ 0.05, I think I will take one more test to check rather than trust this result.

Kruskal-Wallis

```
kruskal.test(drugdt)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  drugdt
## Kruskal-Wallis chi-squared = 3.9865, df = 1, p-value = 0.04587
```

The p-value 0.04587 which is approximately equal to 0.05, so that we fail to reject the null hypothesis and state that the medians are NOT significantly different.

Check the distribution difference between 2 groups:

Kolmogorov-Smirnov:

```
ks.test(drugdt$New.Drug, drugdt$Placebo,alternative = 'greater')
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  drugdt$New.Drug and drugdt$Placebo
## D^+ = 0.6, p-value = 0.1429
## alternative hypothesis: the CDF of x lies above that of y
```

The p-value is 0.1429 bigger than 0.05, so we accept the null hypothesis which mean that the two dataset values are from the same continuous distribution.

Test if variables are correlated

Spearman test

Now, let's check if those 2 variables are correlated together or not with Spearman test:

```
library(pspearman)
spearman.test(drugdt$Placebo, drugdt$New.Drug)
```

```
##
##  Spearman's rank correlation rho
##
## data:  drugdt$Placebo and drugdt$New.Drug
## S = 28, p-value = 0.5167
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##  rho
## -0.4
```

According to the Spearman test:

- The p-value is $0.5167 > 0.05$ which indicates that there's a weak evidence to reject the null Hypothesis. We fail to reject the null hypothesis and there is NO relationship between 2 variables.

Kendall's Tau

```
library(Kendall)
Kendall(drugdt$Placebo, drugdt$New.Drug)
```

```
## tau = -0.4, 2-sided pvalue =0.46243
```

The p-value of this test also bigger than 0.05 which means there's NO relationship between 2 variables.

SUMMARY:

I firstly import the "placebo_new_drug" data which is about results from testing new drug and choose non-parametric test to analyze the data:

- Check the difference between medians of 2 groups with: SIGN test, Wilcoxon,and Kruskal test: the p-values of SIGN and Kruskal test are bigger or equal to 0.05, so the medians are NOT significantly different. There was a red warning when running the Wilcoxon Signed Rank test due to the small size sample, so that I don't trust this result when the p-value = $0.02928 < 0.05$.
- Check the distribution difference: Kolmogorov-Smirnov test results in the p-value is 0.1429 bigger than 0.05, so we can accept the null hypothesis which mean that the two data set values are from the same continuous distribution.
- Check the correlation between 2 variables: Spearman test and Kendal'Tau both result in the conclusion that there's NO relationship between 2 variables.