Non-Parametric Statistics

Ahmad Alqurashi

10/12/2021

Introduction

In week 7, we are going to use Non-Parametric tests to test our dataset. Non-parametric are group of test that typically test central tendency of two populations, usually tests the median because most of study and data collected are not normally distributed. in non-parametric tests there is no assumption about the distribution like Gaussian, t-distribution, or F-distribution. Also, Samples of data are usually small compared to parametric test. Datasets are usually less than 30 observations. Non-parametric testing is used when the underlying distribution is unknown, data is ordinal or ranked, Data contains a lot of outliers or is very skewed, or Small sample sizes.

Types of non-parametric testing:

Difference in medians

- Sign test
- Wilcoxen signed-rank test
- Mann-Whitney U-test
- Kruskal-Wallis test

Difference in distribution

- Kolmogorov-Smirnov test
- Anderson-Darling Test
- Spearman's Rank Correlation Coefficient

Dataset

Dataset that in use for this assignment contains two groups. First group are patients who use placebo drug, the second group are using the actual new drug. data contains total of 5 observations and there is not any missing data.

Method and result

At first, we import required libraries that provide Non-Parametric tests functions.

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

Then, we load dataset into R

```
dt <- fread('C:\\Users\\Ahmad\\Desktop\\MSDS\\MSDS660\\week 7\\assignment\\placebo_new_drug.csv')
```

Here, we calculate median for both group to see if they are significantly different.

```
median(dt$Placebo)
```

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

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

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

Median for Placebo is 6 while the new drug is 3, which they are different.

Fist test we are using is SIGN test. SIGN test is used to show whether the medians of two paired are significantly different or not. It requires two sample of paired data, symmetric distribution of data, dependent variable is ordinal, inerval, or ratio, and independent variable is factor with two level.

H0: there is zero difference between group A and group B's medians.

Ha: there is a difference between group A and group B's medians.

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

```
##
##  Dependent-samples Sign-Test
##
## data:  dt$Placebo and dt$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 is 0.1875 and we fail to reject the null hypothesis. Although, the test shows that true median difference is greater than 0.

Next test is Wilcoxen signed-rank test. This test is the same as SIGN test except it takes magnitude and sign of differences into consideration.

H0: there is zero difference between group A and group B's medians.

Ha: there is a difference between group A and group B's medians.

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

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

p-value = 0.02928, so we reject null hypothesis and stating that median is significantly different.

The test below is Mann-Whitney U-test. It is the same as two previous tests but it Tests if two independent samples come from the same distribution. Meaning that it is two-sided test which in this case it will test whether the new drug is greater or less effective than the placebo.

H0: The medians of values for each group are equal.

Ha: The medians of values for each group are not equal.

```
# two sided

wilcox.test(dt$Placebo, dt$New.Drug)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dt$Placebo and dt$New.Drug
## W = 22, p-value = 0.05855
## alternative hypothesis: true location shift is not equal to 0
```

p-value = 0.05855 so we fail to reject null hypothesis.

This test is Kolmogorov-smirnov test which Tests if the distribution between samples are significantly different.

H0: Distribution between samples are from the same distribution.

H1: Distribution between samples are not from the same distribution.

```
ks.test(dt$Placebo, dt$New.Drug)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  dt$Placebo and dt$New.Drug
## D = 0.6, p-value = 0.3291
## alternative hypothesis: two-sided
```

```r
# confusingly, 'greater' for ks.test means the second set of data's median is greater than the first on
# so if we want the new drug less than the placebo, it's written like this:
ks.test(dt$New.Drug, dt$Placebo, alternative = 'greater')
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  dt$New.Drug and dt$Placebo
## D^+ = 0.6, p-value = 0.1653
## alternative hypothesis: the CDF of x lies above that of y
```

p-value = 0.1653 so we fail to reject null hypothesis. The test shows alternative hypothesis is that the CDF of x lies above that of y.

Lastly, we are going to use spearman correlation test to test if the variables are correlated.

```r
cor(dt[, c('Placebo', 'New.Drug'), with=F], method = 'spearman')
```

```
##          Placebo New.Drug
## Placebo      1.0     -0.4
## New.Drug    -0.4      1.0
```

```r
spearman.test(dt$Placebo, dt$New.Drug)
```

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

p-value = 0.5167 and rho is the Spearman coefficient. the p-value is used to check if there is a strong monotonic relationship between the two variables. So, We accept the null hythpothesis and there is NO relationship between variables.

Conclusion

In sum, we used 4 Non-parametric tests to test the differences between medians and distributions between two paired groups. Other test are not suitable for this dataset for different reasons. For instance, the dataset is too small for Anderson-Darling test because this test needs at least 7 observations to work properly.

References:

Nonparametric Statistics. (n.d.). Retrieved from From the Expert. Non-Parametric Statistics. (2021). Retrieved from ppt .