Nonparametric Statistics

Anisha Mittal

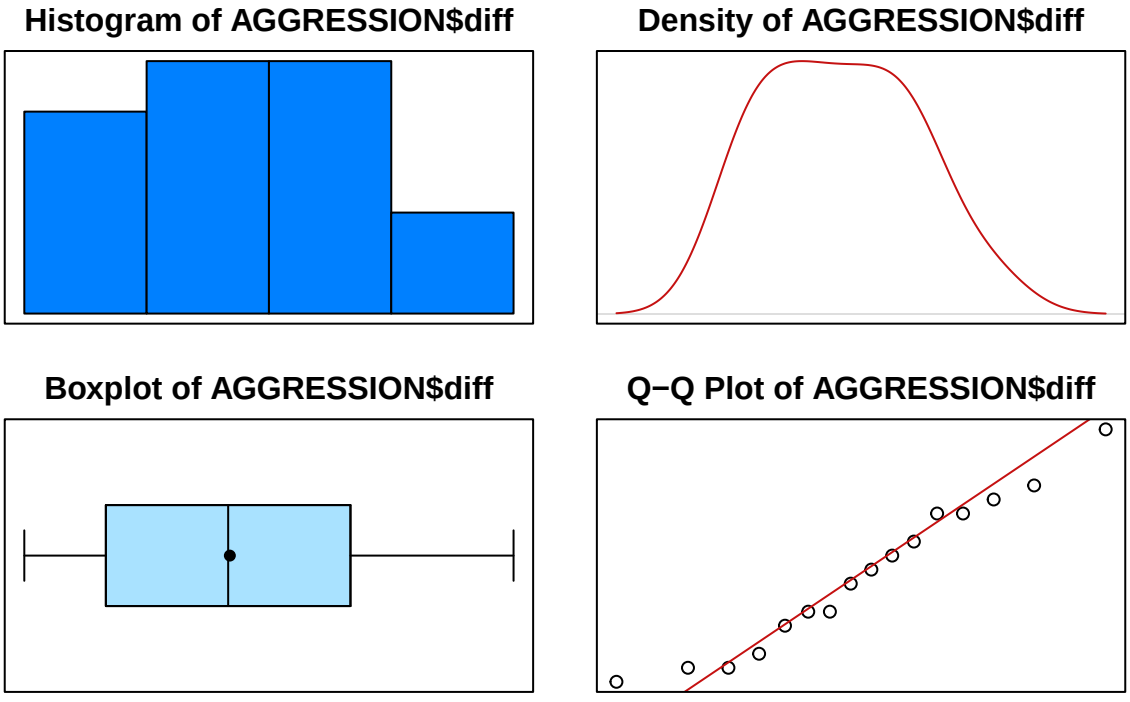
2021-2-2

```
# Loading the package for 'Probability and Statistics with R'
library(PASWR2)
```

a) Loading the dataset AGGRESSION (available in PASWR2 package)

```
data("AGGRESSION") # Loading the dataset
diff = (AGGRESSION$violence - AGGRESSION$noviolence) #Differencing the variables
AGGRESSION$diff = diff # Adding the difference column to the dataset
eda(AGGRESSION$diff) # Exploratory Data Analysis
```

EXPLORATORY DATA ANALYSIS




```
## Size (n)  Missing  Minimum   1st Qu     Mean   Median   TrMean   3rd Qu
```

```
##   16.000    0.000   -3.000    0.500    4.562    4.500    4.562    9.000
##      Max    Stdev      Var  SE Mean   I.Q.R.    Range Kurtosis Skewness
##   15.000    5.341   28.529    1.335    8.500   18.000   -1.199    0.208
## SW p-val
##    0.645
```

By the EDA command , we can see mean = median = 4.5, the density plot is a symmetric bell shaped curve, the boxplot also shows the mean circle on the median line, and the Q-Q plot also shows symmetry. Therefore the distribution of differences is symmetric.

b) Testing whether the median difference > 0 at alpha = 0.05

Here we are using The Wilcoxon signed-rank test with Null Hypothesis: Median = 0 ; Alternative Hypothesis: Median > 0 ; Significance level (alpha): 0.05 ; We reject the null hypothesis if the p-value < alpha

```
# Using wilcoxon signed-rank test and removing ties
wilcoxe.test(AGGRESSION$diff, mu= 0, alternative = "greater")
```

```
##
##  Wilcoxon Signed Rank Test
##
## data:  AGGRESSION$diff
## t+ = 118.5, p-value = 0.003265
## alternative hypothesis: true median is greater than 0
## 95.20569 percent confidence interval:
##    2 Inf
## sample estimates:
## (pseudo)median
##            4.5
```

The Wilcoxon signed-rank test gives test statistic as 118.5 and the p value = 0.003265 < 0.05 (alpha). Therefore we reject the null hypothesis.

c) Creating a Confidence coefficient at 95% level

```
# Using wilcoxon signed-rank test, removing ties and Confidence level is 95%
wilcoxe.test(AGGRESSION$diff , conf.level = 0.95)
```

```
##
##  Wilcoxon Signed Rank Test
##
## data:  AGGRESSION$diff
## t+ = 118.5, p-value = 0.006531
## alternative hypothesis: true median is not equal to 0
## 95.10803 percent confidence interval:
##  1.5 7.5
## sample estimates:
## (pseudo)median
##            4.5
```

d) Repeat b) and c) using The One-sample sign test

Here we are using The One-sample Sign test with Null Hypothesis: Median = 0 ; Alternative Hypothesis: Median > 0 ; Significance level (alpha): 0.05 ; We reject the null hypothesis if the p-value < alpha.

```
# (b) Testing whether the median difference > 0 at alpha = 0.05
with(data = AGGRESSION,SIGN.test(diff,md=0, alternative = "greater"))
```

```
##
##  One-sample Sign-Test
##
## data:  diff
## s = 12, p-value = 0.03841
## alternative hypothesis: true median is greater than 0
## 95 percent confidence interval:
##  1.173947      Inf
## sample estimates:
## median of x
##         4.5
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.8949 2.0000    Inf
## Interpolated CI       0.9500 1.1739    Inf
## Upper Achieved CI     0.9616 1.0000    Inf
```

```
# (c) Creating a Confidence coefficient at 95% level
with(data = AGGRESSION,SIGN.test(diff,md=0,conf.level = 0.95))
```

```
##
##  One-sample Sign-Test
##
## data:  diff
## s = 12, p-value = 0.07681
## alternative hypothesis: true median is not equal to 0
## 95 percent confidence interval:
##  0.03450549 9.00000000
## sample estimates:
## median of x
##         4.5
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level  L.E.pt U.E.pt
## Lower Achieved CI     0.9232  1.0000      9
## Interpolated CI       0.9500  0.0345      9
## Upper Achieved CI     0.9787 -1.0000      9
```

The One-Sample Sign test gives test statistic as 12 and the p value = 0.03841 < 0.05 (alpha). Therefore we reject the null hypothesis.

e) Repeat b) and c) using The t-test

Since we proved that mean = median, that is the distribution of difference is symmetric, we can use the t-test.

Here we are using The One Sample t-test with Null Hypothesis: Median = 0 ; Alternative Hypothesis: Median > 0 ; Significance level (alpha): 0.05 ; We reject the null hypothesis if the p-value < alpha.

```
# (b) Testing whether the median difference > 0 at alpha = 0.05
t.test(AGGRESSION$diff , mu = 0, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  AGGRESSION$diff
## t = 3.4168, df = 15, p-value = 0.001912
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  2.221621      Inf
## sample estimates:
## mean of x
##    4.5625
```

```
# (c) Creating a Confidence coefficient at 95% level
t.test(AGGRESSION$diff , mu = 0)
```

```
##
##  One Sample t-test
##
## data:  AGGRESSION$diff
## t = 3.4168, df = 15, p-value = 0.003824
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  1.716338 7.408662
## sample estimates:
## mean of x
##    4.5625
```

The One-Sample t-test gives test statistic as 3.4168 the p value = 0.001912 < 0.05 (alpha). Therefore we reject the null hypothesis.

The one sample t-test gives the least p-value as compared to the other two tests which shows it is more accurate, therefore I would prefer t-test.