Non-parametric Tests 2

1. Objectives

- Explain the R procedures to conduct and check assumptions for the Sign test and Wilcoxon signed-rank sum test.
- Understand and interpret the R output.

2. Procedures

The Wilcoxon signed-rank sum test is a **non-parametric** approach used to compare paired data when the data are not normally distributed. For Wilcoxon signed rank sum test, just follow the same code as Mann-Whitney-Wilcoxon (MWW) test except for the argument `paired=TRUE`:

```
➢ wilcox.test(x,y,alternative="two.sided",paired=TRUE,exact=NULL,
  correct=TRUE)# data are saved in two different numeric vectors
➢ wilcox.test(outcome ~ grouping variable, data = name of data frame,
  alternative = "two.sided", paired = TRUE)# data are saved in a data
  frame
```

For the Sign test, we must install the package **`PASWR`**, so that we can use function **`SIGN.test`**:

```
➢ install.packages("PASWR")
➢ library(PASWR)
➢ SIGN.test(x,y,alternative="two.sided")
```

Don't forget to check the assumptions to decide which test to use. We prefer Sign test for ranked data, and Wilcoxon signed-rank sum test for quantitative data (with non-normal differences).

3. Exercises

Exercise 1. A test was conducted for two overnight mail delivery services. Two samples of identical deliveries were set up so that both delivery services were notified of the need for a delivery at the same time. The hours required to make each delivery are stored in **Overnight** data file. Do the data suggest a difference in the delivery times for the two services? Use a 0.05 level of significance for the test.

We're going to work with the file **Overnight.csv**, so firstly import it into R:

```
➢ overnight<-read.table("Overnight.csv", header  = T, sep=",",
  stringsAsFactors=F)
➢ head(overnight) #to see some first subjects in the dataframe
➢ str(overnight) #to see the structure of the dataframe
```

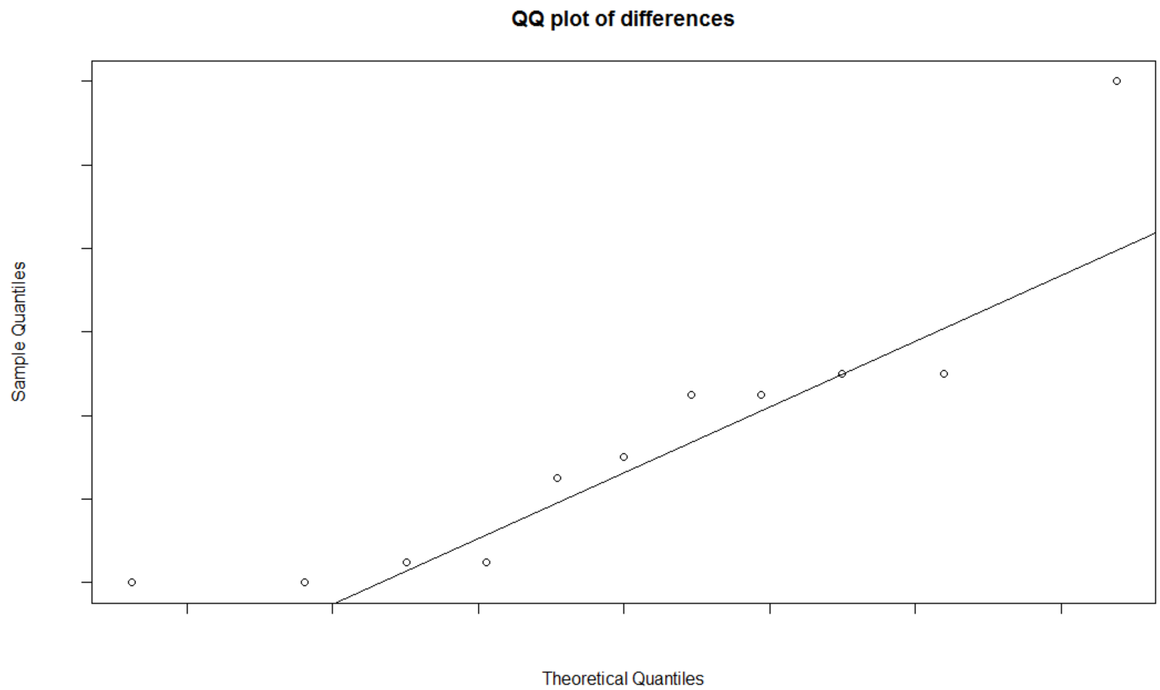
The next step is to check the assumptions to see which test is to be applicable in this case. From the above codes, we know that the data are quantitative and the two samples are matched. Let's check the normality of differences with the help of the stem-and-leaf display and the Q-Q plot.

```
➢ diff<-overnight$Service1-overnight$Service2
➢ stem(diff)
➢ qqnorm(diff,main="QQ plot of differences")
➢ qqline(diff)
```

The outputs are given below.

```
The decimal point is 1 digit(s) to the right of the |

  -0 | 444421
   0 | 1111
   0 | 8
```



Based on the R outputs, it's not reasonable to assume that the differences in delivery times are normally distributed. As a result, we must use a nonparametric test instead of a parametric one (t-test for matched pairs). In more details, ***we're comparing 2 populations; samples are matched; data are quantitative but differences between paired samples cannot be assumed to be normal***, so we must apply the Wilcoxon signed-rank sum test.

Question 1. Set up the hypotheses for the test. What type of test is this?

Now we apply `wilcox.test()` to produce the R output for this problem. Notice that we are asked to test for a significant difference between the 2 groups, choose `alternative="two.sided"`; and with paired samples, we set `paired=TRUE`.

```
➤ ex1<-wilcox.test(overnight$Service1,overnight$Service2,alternative
  ="two.sided",paired = TRUE,correct = TRUE)
➤ ex1

        Wilcoxon signed rank test with continuity correction

data:  overnight$Service1 and overnight$Service2
V = 22, p-value = 0.3489
alternative hypothesis: true location shift is not equal to 0
```

Question 2. What is your conclusion in this case?

Exercise 2. Vendors of prepared food are very sensitive to the public's perception of the safety of the food they sell. Food sold at outdoor fairs and festivals may be less safe than food sold in restaurants because it is prepared in temporary locations and often by volunteer help. What do people who attend fairs think about the safety of the food served? One study asked this question of people at a number of fairs in the Midwest: How often do you think people become sick because the food they consume are prepared at outdoor fairs and festivals? The variable "**sfair**" contains the responses described in the example concerning safety of food served at outdoor fairs and festivals. The variable "**srest**" contains responses to the same question asked about food served in restaurants. The possible responses were: 1 = very rarely; 2 = once in a while; 3 = often; 4 = more often than not; and 5 = always. In all, 303 people answered the question. We suspect that restaurant food will appear safer than food served outdoors at a fair. Do the data give good evidence for this suspicion? Conduct the approporiate test with significance level $\alpha = 0.05$.

The data are stored on file **foodsafety.csv**, so import it into R and check some first subjects as follows.

```
  subject hfair sfair sfast srest gender
1       1     4     1     1     1      1
2       2     4     2     4     2      1
3       3     2     2     2     2      1
4       4     4     2     2     1      1
5       5     2     3     1     3      1
6       6     1     2     2     2      1
```

Check the assumptions and choose the appropriate technique.

Question 3. Are the two samples independent or matched? What type of data are represented? Should we conduct the parametric test in this case?

On answering the above questions, you would realize that the best choice is the Sign test. Using `SIGN.test()` gives you the following results:

- `ex2<-SIGN.test(foodsafety$sfair,foodsafety$srest,alternative="greater)`
- `ex2`

```
        Dependent-samples Sign-Test

data:  foodsafety$sfair and foodsafety$srest
S = 137, p-value < 2.2e-16
alternative hypothesis: true median difference is greater than 0
95 percent confidence interval:
   0 Inf
sample estimates:
median of x-y
            0
```

Question 4. Give your interpretation of this output.

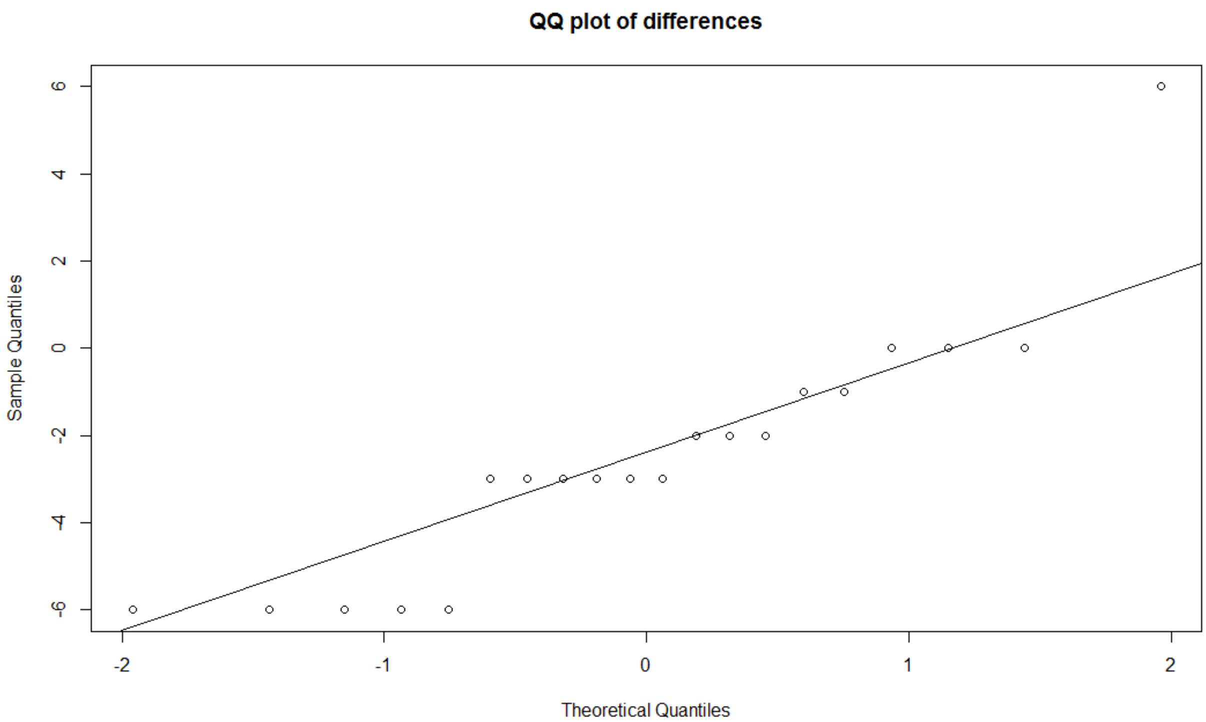
Exercise 3. File **french.csv** presents the scores in a test of understanding of spoken French for a group of executives before and after an intensive French course.

(a) Show the assignment of ranks and the calculation of the signed rank statistic T^+ for the test to see that the mean improvement in scores before and after the course is different than 0.

(b) Now use R to implement the Wilcoxon signed-rank procedure to reach a conclusion about the impact of language course. State the hypotheses in words and report the statistic T^+, its p-value, and your conclusion. Remember to check all the assumptions.

Here are the outputs.

```
  Executive Pretest Posttest
1         1      32       34
2         2      31       31
3         3      29       35
4         4      10       16
5         5      30       33
6         6      33       36
```




```
  The decimal point is 1 digit(s) to the right of the |

  -0 | 66666
  -0 | 33333322211
   0 | 000
   0 | 6

	Wilcoxon signed rank test with continuity correction

data:  french$Pretest and french$Posttest
V = 14.5, p-value = 0.003257
alternative hypothesis: true location shift is not equal to 0
```

Exercise 4. A student organization surveyed both current students and recent graduates to obtain information on the quality of teaching at a particular university. An analysis of the responses provided the following teaching-ability rankings stored in **Professors.csv**. Do the rankings given by the current students agree with the rankings given by the recent graduates? Use $\alpha = 0.1$ to draw conclusion.

```
  Professor Current.Students Recent.Graduates
1         1                4                6
2         2                6                8
3         3                8                5
4         4                3                1
5         5                1                2
6         6                2                3

'data.frame': 10 obs. of  3 variables:
 $ Professor       : int  1 2 3 4 5 6 7 8 9 10
 $ Current.Students: int  4 6 8 3 1 2 5 10 7 9
 $ Recent.Graduates: int  6 8 5 1 2 3 7 9 4 10


       Dependent-samples Sign-Test

data:  professors$Current.Students and professors$Recent.Graduates
S = 4, p-value = 0.7539
alternative hypothesis: true median difference is not equal to 0
95 percent confidence interval:
 -2.000000  2.675556
sample estimates:
median of x-y
           -1
```