Lab 5

Q.1:
You want to see if the mean of goals suffered by two football teams over the years is the same. The number of goals suffered by each team in 6 games for each year are as follows:

Team X: 6, 8, 2, 4, 4, 5
Team Y: 7, 10, 4, 3, 5, 6
($\alpha = 0.05$)

(**Hint: Wilcoxon rank sum test** is used when is asked to compare the means of two groups that do not follow a normal distribution: it is a non-parametrical test. It is the equivalent of the t test, applied for independent samples)

Q.2:
The manager of a national park wants to see if pollution levels in lake are reduced by banning the boat rental. This is measured by the rate of pollution every 60 minutes (8am 22pm: total of 15 measurements) in a day when rental is allowed, and in a day when it is not. Here are the values of water pollution:

With rental: 214, 159, 169, 202, 103, 119, 200, 109, 132, 142, 194, 104, 219, 119, 234
Without rental: 159, 135, 141, 101, 102, 168, 62, 167, 174, 159, 66, 118, 181, 171, 112
($\alpha = 0.05$)

A) Are these two groups of sampling paired or independent? Why?
B) Use Wilcoxon signed rank test to calculate p-value. Based on the mentioned significant level ($\alpha = 0.05$), what do you conclude about banning the boat rental for any improvements in terms of pollution of the lake?
C) Calculate the sum of ranks assigned to the differences with positive sign, and the sum of ranks assigned to the differences with negative sign:
1. Calculate the vector containing the differences (diff)
2. Delete all differences equal to zero
3. Check the ranks of the differences, taken in absolute, recalling the signs of the values of the differences (Calculate diff.rank vector)
4. Check the sign to the ranks, recalling the signs of the values of the differences (Calculate diff.rank.sign vector)
5. Calculate the sum of ranks assigned to the differences as a positive (ranks.pos).
6. Calculate the sum of ranks assigned to the differences as a negative (ranks.neg)

Q.3:

Suppose we are going to compare two software **Zip** and **Tar** on archiving 20 different files. The execution time (time from the beginning till the end of the archiving files in milliseconds) is recorded as follows:

Zip execution time: 10, 44, 65, 77, 43, 44, 22, 66, 50, 100, 55, 99, 44, 23, 100, 88, 200, 220, 110, 551
Tar execution time: 20, 55, 75, 60, 55, 88, 35, 33, 35, 80, 65, 82, 47, 35, 97, 110, 250, 190, 111, 600

Your task is to find out if there is any significant difference in the execution times of two different archiving programs based on their running time on same file.

A) Are these two groups of sampling paired or independent? Why?

B) Plot the distribution of these two groups of sampling to see if they have normal distribution or not.

C) Use Wilcoxon signed rank test to calculate p-value. Based on the mentioned significant level (α = 0.05), what do you conclude about differences between the mean of run-time of two different archiving programs?