EXERCISES FOR STATISTICAL TESTS BASED ON RANKS

Exercise 4.1. Write a function to perform the Wilcoxon Rank-Sum test for a difference in medians. The preamble of this function should be as follows:

```
diff.mean.nonparametric = function(data1, data2, alpha=0.05) {
1
   ### PERFORMS MANN-WHITNEY TEST FOR A DIFFERENCE IN MEDIANS
2
   ### INPUT:
   ###
         DATA1[N1]: SAMPLE 1
         DATA2[N2]: SAMPLE 2
   ###
5
   ###
         ALPHA: SIGNIFICANCE LEVEL (DEFAULT = 5%)
   ### OUTPUT: LIST$
   ###
         Z: Z-VALUE FOR THE NORMAL APPROXIMATION TO MANN-WHITNEY TEST
         Z.CRIT: CRITICAL VALUE FOR ALPHA*100% SIGNIFICANCE
   ###
10
   ###
         PVAL: P-VALUE OF THE TEST
```

The following R functions are useful for non-parametric testing: rank(), sort(), order(). You can read the documentation for these functions and play with them by generating the random numbers seeing what happens:

Exercises for Statistical Methods for Climate Scientists. By DelSole and Tippett

14 EXERCISES FOR STATISTICAL TESTS BASED ON RANKS

```
> x = rnorm(5)
  > x
2
   [1] 0.5726722 -0.6338714 0.3260635 -1.6519467 0.9540662
  > sort(x)
   [1] -1.6519467 -0.6338714 0.3260635 0.5726722 0.9540662
  > order(x)
6
   [1] 4 2 3 1 5
7
  > x[order(x)]
   [1] -1.6519467 -0.6338714 0.3260635 0.5726722 0.9540662
  > rank(x)
10
   [1] 4 2 3 1 5
11
```

Rank is a very useful function for this homework set because it automatically deals with ties correctly.

```
1 > x[1] = x[2]
2 > x
3 [1] -0.6338714 -0.6338714 0.3260635 -1.6519467 0.9540662
4 > rank(x)
5 [1] 2.5 2.5 4.0 1.0 5.0
```

Exercise 4.2. Run this function on exactly the same PDO data you used in the previous 2 exercise sets to test if the median before and after 1977 differ. Clearly state the z-value, critical z value, and the p-value of the test. Clearly state whether you reject the null hypothesis or not, and whether this conclusion is consistent with your conclusions on previous homeworks. You can check your answer using wilcox.test(pdo1,pdo2)

Exercise 4.3. In exercise 2.1, you wrote the function cor.equal.test to test the correlation coefficient. Write a new function that performs a correlation test based on Spearman's rank correlation. The preamble of this function should look like the following:

```
cor.test.nonparametric = function(data1, data2, alpha=0.05) {
   ### PERFORMS RANK CORRELATION TEST
2
   ### INPUT:
3
   ###
         DATA1[N1]: SAMPLE 1
4
   ###
         DATA2[N2]: SAMPLE 2
5
         ALPHA: SIGNIFICANCE LEVEL (DEFAULT = 5%)
   ###
   ### OUTPUT: LIST$
         RHO.SPEARMAN = SPEARMAN'S RANK CORRELATION
   ###
   ###
         ZVAL: Z-VALUE FOR THE NORMAL APPROXIMATION TO MANN-WHITNEY TEST
   ###
         Z.CRIT: CRITICAL VALUE FOR ALPHA*100% SIGNIFICANCE
10
   ###
         PVAL: P-VALUE OF THE TEST
11
```

Exercise 4.4. Run this function on the same PDO data you used in previous exercises to test if the lag-1 correlation is significant before 1977, after 1977, and the full period. State whether the procedure rejects the null hypothesis or not. Comment on whether the result

is consistent with the hypothesis tests you performed in exercise 2.1. You can check your answer to some extent using

cor.test(pdo1[10],pdo1[10+1],method='spearman',continuity=FALSE,exact=FALSE)

However, the p-value will not be exactly the same (but they are close). The reason they are not the same is because R computes p-values using

$$t = \frac{\hat{\rho}_{Sp}\sqrt{N-2}}{\sqrt{1-\hat{\rho}_{Sp}^2}},$$
(4.1)

and then it refers to a t-distribution with N-2 degrees of freedom.

Exercise 4.5. Write a function called diff.dispersion for testing a difference in dispersion. The preamble of this function should be as follows:

```
diff.dispersion = function(data1, data2, alpha=0.05) {
1
   ## PERFORMS WILCOXON SOUARED-RANK TEST
2
   ## ON ABSOLUTE DEVIATIONS |X - MEDIAN(X)|
3
   ### INPUT:
4
   ###
        DATA1[N1]: SAMPLE 1
   ###
         DATA2[N2]: SAMPLE 2
   ###
        ALPHA: SIGNIFICANCE LEVEL (DEFAULT = 5%)
7
   ### OUTPUT: LIST$
   ###
         ZVAL: Z-VALUE FOR THE NORMAL APPROXIMATION TO MANN-WHITNEY TEST
         Z.CRIT: CRITICAL VALUE FOR ALPHA*100% SIGNIFICANCE
   ###
10
         PVAL: P-VALUE OF THE TEST
11
   ###
```

Run your function on exactly the same PDO data you used in previous exercises to test if the variance before and after 1977 differ. Clearly state the z-value, critical z value, and the p-value of the test. Clearly state whether you reject the null hypothesis or not, and whether this conclusion is consistent with your conclusions on previous homeworks. (Hint: most of this function is similar to your difference in medians test.)