ST204 Nonparametric Statistics
2021-22 Semester 2
Assignment Sheet 1

Due at 16:00 on Friday 25th February 2022.

Only one randomly chosen question will be marked. Submit answers to questions not already covered in tutorials only. Your submission file should be in the form of a merged .pdf and your code must be provided for any questions involving R. Otherwise, you are free to mix typed/handwritten solutions as you see fit.

1. Consider the dataset: 1.2, -0.3, 4.6, 7.8, 2.1.

   Is there evidence that the population median θ differs from 0? Test this using the Wilcoxon signed-rank test. Compute the test statistic by hand (i.e. show your workings) and find the p-value using R. Give the conclusions of the test.

2. The annual rainfall (cm) was recorded in a city for 34 consecutive years. The values are:
   54, 73, 75, 58, 69, 72, 63, 72, 66, 57, 69, 84, 73, 63, 62, 69, 78, 68, 48, 49, 71, 75, 76, 70, 70, 82, 71, 78, 51, 51, 85, 67, 78, 84.

   It is hypothesised that the true population median is 69.

   (a) For this data, would using the sign test or the Wilcoxon signed-rank test approach be more appropriate for constructing a confidence interval for the median? You may use R as a guide.

   (b) Using R, construct a confidence interval with *at least* 99% coverage for the median using the most appropriate method identified in (a). Interpret the CI and state the exact confidence level achieved.

   (c) Are there any potential problems with the computed confidence interval for this data?

3. The survival times in weeks of ten patients suffering from a serious disease were recorded during a medical study as: 49, 58, 75, 110, 112, 132, 151, 276, 281, 362.

   Answer the questions below by hand, using the statistical tables provided on Moodle where necessary.

   (a) Use a sign test to test the hypothesis that the median survival time, θ, differs from 200. You must calculate the p-value by hand. State your conclusion.

   (b) Repeat the above test using the normal approximation to the sign test. Do you reach the same conclusion?

4. The average energy expenditures for eight elderly women were estimated on the basis of information received from a battery powered heart rate monitor that each subject wore. Two overall averages were calculated for each woman, one for the summer months and one for the winter months. Here is the dataset:

Subject, i	1	2	3	4	5	6	7	8
Summer, x_i	1458	1553	2209	1804	1912	1366	1686	1556
Winter, y_i	1424	1103	1495	1739	2031	934	1401	1339

   You *must* use R to answer the questions below.

   (a) Is there evidence of a difference in energy consumption between the two seasons, according to the Wilcoxon signed-rank test?

   (b) Give **one** reason why you think such a test is appropriate to use for these data.

   (c) Verify that $[65, 441]$ is an approximate 90% confidence interval for the median population difference using the normal approximation to the Wilcoxon distribution, this time *without using the continuity correction.*