## CS544 Module 2 Assignment

**General Rules for Homework Assignments**

* You must work on your assignments individually. You are not allowed to copy the answers from the others.
* Each assignment has a strict deadline. If there is a delay, you must be in touch with the instructor. Late submissions without reasons will result in a grade deduction.

15%

* When the term *lastName* is referenced in an assignment, please replace it with your last name.

# Part1) Probability - 60 points

For questions a) and b), show the solutions with the calculations without using R. Then, verify with R code.

1. A disease affects 5 out of 100 people on average. The sensitivity of a clinical test to detect the disease is 92%, which means 92% of people who have the disease get positive test results. Its false positive rate is 6%, which means 6% of people who do not have the disease get positive results in the tests. What is the chance that a randomly selected person with a positive result does not have the disease? What is the chance that a randomly selected person with a negative result actually has the disease?
2. Suppose that in a particular state, among the registered voters, 42% are Democrats, 48% are Republicans, and the rest are independents. A ballot question is whether to provide universal healthcare to citizens. Suppose that 85% of Democrats, 50% of Republicans, and 60% of Independents favor universal healthcare. If a person chosen at random **does not** favor universal healthcare, what is the probability that the person is i) a Democrat? ii) a Republican, iii) an Independent.
3. In a drama class, there are 12 females and 10 males. A group of 15 is to be chosen at random for a play. i) What is the probability that all males were chosen? ii) What is the probability that at least 5 males were chosen? Iii) What is the probability that at most 5 males were chosen?

# Part2) R - 40 points

Using function data() to load R data set *airquality*.

Provide R code and output for all of the following.

1. Use the diff function to calculate the temperature differences between consecutive days.

Insert the value 0 at the beginning of these differences. Add this result as the DIFFS column of the data frame.

1. Calculate the number of days that are warmer than the previous day?
2. Show the mean and median temperature for each month (May to Sept.). Do not hard code the month. Print out your result properly. For example, “The average and median temperature for May are xxx and xxx; …”.
3. Show the coldest and hottest days each month (May to Sept.). Do not hard code the month. Print out your result properly. For example, “The coldest and hottest days in May are xxx and xxx; …”.

# Submission:

Upload your result file to the Assignments section of Blackboard.

Provide all R code in a single file, CS544\_lastName.R. Clearly mark each subpart of each question and add appropriate comments.

If you need to submit more than one file, create a folder, CS544\_lastName, and place all files in this folder. Archive the folder (CS544\_A1\_lastName.zip).