Homework Assignment 4
(Due Friday October 11, 2024 before class)
Total Points: 74

Please hand in a hard copy of your homework (compiled pdf file from R markdown) in class and email your R code to Kaniz Fatema (KFATEMA@email.sc.edu). Please use the R markdown Homework template (HWtemplate.Rmd) to write your homework solutions. Work on the homework independently.

Problem 1. Use the ALL dataset and create the following plot. (15 points) [Hint: Use the R code in http://people.stat.sc.edu/hoyen/STAT504/ALLcode.R to access the expression data. stripchart(y~x, method= "jitter", jitter=0.2, vertical=T, ylab=…, main=…) where y is the expression data and x is the cancer molecular subtypes (mol.biol). Use ylab to label y-axis correctly and main to create a main title. For the blue lines indicating means in each group, use lines(c(x1,x2), c(y1,y2), col=4) where x1, x2, y1, y2 are the locations of the line in x-axis and y-axis respectively.]

Distribution of 1636_g_at probe by cancer molecular subtypes
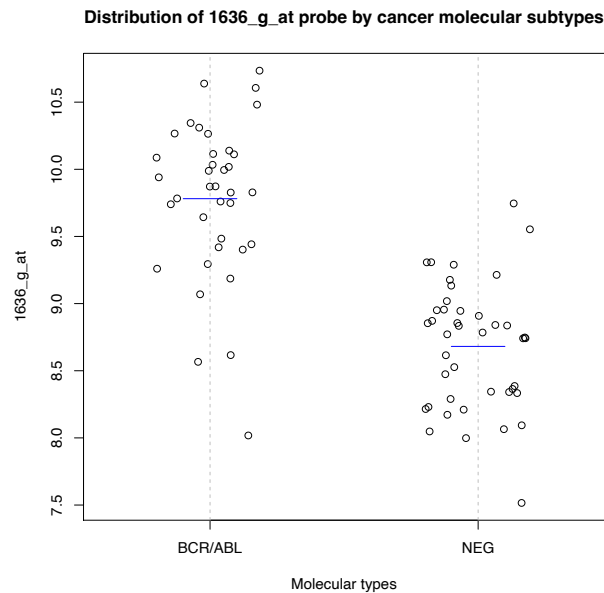
Problem 2. This exercise is for practicing central limit theorem.

(a) Draw n=5 samples from uniform distribution and calculate sample means. Repeat this experiment 200 times, plot the distribution of sample means. (10 points) [Hint: To simulate n samples from uniform distribution, use runif(n). Use plot(density(x)), where x is the vector contains the sample means from these 200 experiments.]

(b) Repeat (a) but use n=100 (3 points).

(c) Compare the sample distributions obtained in (a) and (b), what do you observe? (10 points)

Problem 3. Perform the following steps in R:

(a) Simulate 30 samples from Normal(mean=0, sd=1) (2 points)

(b) Randomly assign 15 samples into control and 15 into treatment group (15 points) [Hint: Use sample]

(c) Perform two sample T-test and report the p value. (2 points)

(d) Randomly generate 1000 samples from uniform distribution, and plot the histogram of the 1000 samples. [Hint: Use hist(x) to plot a histogram of x.] (2 points)

(e) Repeat (a) (b) (c) 1000 times, and stored the corresponding 1000 p values in a vector, plot a histogram using these 1000 p values. What is the distribution of p values? (15 points)