ST437 / ST537
Assignment 2
(Please include relevant R code and output)

1. Consider the wastewater data example ("Wastewater.txt": http://www4.stat.ncsu.edu/~staicu/Spring2018/data/)
   a. The objective is to investigate the reliability of the data produced by the commercial lab relative to the Wisconsin State Laboratory of Hygiene. Make sure you validate all the assumptions you make.
   b. Construct a 99% confidence region for the mean difference of biochemical oxygen demand (BOD) and suspended solids (SS).
   c. Construct joint confidence intervals for BOD and SS using the simultaneous method as well as the Bonferroni correction approach.
   d. State your results.

ST437: only (a)-(c)

Questions 2 and 3 are based on the data frame bottle.df in the Hotelling library; the data summarizes a study of the chemical composition of six different Heineken beer bottles. Each bottle was examined twenty times, and the glass was assayed for the composition of the elements manganese, barium, strontium, zirconium, and titanium.

To read the data:
   library(Hotelling)
   data(bottle.df)
   bottle.df.

2. Consider the data for the first two bottles only.
   a. Examine the data using plots to identify outliers. Identify any remarkable outliers or abnormalities. Test the data for multivariate normal distribution, both marginally and within each of the two bottles.
   b. Apply the Hotelling test to examine whether the multivariate normal means are the same for the two bottles.
   c. Give the 95% confidence region for the mean difference.

ST437: (a)-(b)

3. Consider the full data
   a. Examine the data using plots to identify outliers. Identify any remarkable outliers or abnormalities. Test the data for multivariate normal distribution for all the bottles.
   b. Examine whether the multivariate normal means are the same across the six bottles.
   c. Do a pairwise comparison to determine which bottles or chemical decompositions is
different using 95% confidence intervals that have simultaneous interpretation.

ST437: (a)-(c)

4. Observations on two responses (called Resp1, Resp2) are collected for three treatments; see Example_0125.txt. Data are at http://www4.stat.ncsu.edu/~staicu/Spring2018/data/

a. Write down the one-way MANOVA model for this data and provide estimates for the overall mean and the treatment effect. As well write down the residuals.

b. Construct the one-way MANOVA table by calculating the matrices that go into each table

c. Evaluate the Wilks’ lambda for level of significance alpha=0.05

ST437: (a)-(c)

5. The data (“skull_size.dat”) can be found at: http://www4.stat.ncsu.edu/~staicu/Spring2018/data/

Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations. Four measurements were made of male Egyptian skulls for three different periods: period 1 (4000 BC), period 2 (3300 BC) and period 3 (1850 BC). The measured variables are:

X1 - maximum breadth of skull (mm)
X2 – basibregmatic height of skull (mm)
X3 – basialveolar length of skull (mm)
X4 – nasal height of skull (mm)

In the data, the first variable is denoted by V1, the second by V2 and so on. The last variable, denoted by V5 corresponds to the period time.

Construct a one-way MANOVA of the Egyptian skull data. Use alpha=0.05. Construct 95% simultaneous confidence intervals to determine which components differ among the populations presented for the three time periods. Interpret the results. In your response make sure you validate all the assumptions being made.