STAT 4043/ 5543

Assignment 1 (25 points + 2 bonus points)  
Due: By 10:30am on 9/8/2023

Please answer the following questions briefly. You can create extra space below each question to answer there or answer in separate sheets. Both typed-in answers and handwritten answers are accepted. Electronic submission (Canvas) only.

**Please submit a single file** and **include your name in the body of the file** (not filename).

**You must submit your R code** (copy and paste at the end of your electronic file as an appendix). If the R code is not submitted, 50% of the points will be deducted on each question that asked you to use R. Please use comments in R to indicate which part of the code corresponds to which question. If a question does not explicitly ask you to use R, then it is your choice whether you want to use R or do it by hand. **You must include all the plots asked by the questions in your answer.**

1. You should solve this problem by hand without using R.

Consider the three matrices **C** = .

1. Is **B** a symmetric matrix? Write yes or no.
2. Is **C** a diagonal matrix? Write yes or no.
3. Find the transpose of **A**.
4. Define the 3 matrices in R and find . Your code will also be graded for this problem.

(1+ 1 + 1 + 2= 5 points)

1. A criminologist studying the relationship between the level of education and crime rate in medium-sized U.S counties collected the data given in the file ‘Crime rate.txt’ for a random sample of 84 counties; the second column is the percentage of the individuals in the county having at least a high-school diploma, and the first column is the crime rate (crimes reported per 100,000 residents) last year. The criminologist wants to build a linear model to see if the crime rate depends on the percentage of high school diplomas.
2. Read the data into R using the function read.table(). This works exactly like read.csv(), but it can read data from files that are not necessarily a csv file. Make sure you use the appropriate value of the header argument (TRUE or FALSE). (1 points)
3. An expert decided that the observations with less 80% rate of high school diploma may not be reliable. So, we will only use the part of the data for which this rate is at least 80%. Use R to find this subset of the data. How many observations does it have? Find the average crime rate from this data set. (3 points)
4. Use this subset of the data to obtain the correlation between crime rate and high school diploma percentage. Also use a hypothesis testing method to test if the correlation is significantly different from zero at 1% level. You must write the null and the alternative hypotheses, justify your reason to accept/reject the null hypothesis at a pre-specified level, and write your final conclusion. (1 + 4 points)
5. Use the original data set (not the subset obtained in part (b)) to answer this question and the following questions. Create a scatter plot of the two variables with the crime rate in the Y-axis. Make the points green solid dots and make their size 80% of the default. Put appropriate axis labels. The axis labels should be 75% of the default size. Add an appropriate title.   
   Then use the **points ()** function to make the data points coming from counties with less 80% high school diploma rate orange. You must include the plot in your answer. (6 points)
6. Find the average crime rate in the counties with less than 80% of a high-school diploma. Find the same average for counties with more than or equal to 80% of high school diplomas. (2 points)
7. A contract engineer at DuPont Corp. studied the rate at which a spilled volatile liquid will spread across the surface (Chemical Engineering Progress, January 2005). Assume 50 gallons of methanol spills onto a level surface outdoors. The engineer used derived empirical formulas (assuming a state of turbulent free convection) to calculate the mass (in pounds) of the spill after a period of time ranging from 0 to 60 minutes. You will find the data in the file LIQUIDSPILL.Rdata.
8. Create an appropriate scatter plot with the time in the horizontal axis and mass in the vertical axis. Use appropriate axis labels.
9. Based on the plot, does it seem like there is a relationship between the two variables? If yes, is it a linear relationship?   
     
   (3 + 2 bonus)