STAT 5385/ STAT 6385: Project 1 (Due on 11.30 pm Friday, September 15, by email)
Total Marks 50

Notes:

- For starting seed value, use the last 4 digits of your ID.
- You are supposed to work on this project entirely on your own. So, do not consult with anyone within or outside the class.
- You are welcome to ask me questions. However, first try to find the answer on your own. Don't be afraid to google! Google is the best friend of a graduate student!!

1. The algae dataset contains data on 90 independent river water samples. For each water sample, the following variables were recorded:

Table 1: Overview of the datasets.

#	Variable	Description
1	season	the season of the year (1, 2, 3, and 4 for four seasons)
2	river.size	size of the river (1, 2, and 3 for small, medium, and large)
3	fluid velocity	fluid velocity (1, 2, and 3 for low, medium, and high)
4-11	chem #	concentration of eight chemical substances
12	abundance	abundance of a certain class of algae (in log scale)

   (a) Is river size associated with fluid velocity? Carry out an appropriate test. Include the appropriate hypotheses, test statistic value, p-value, and conclusion. [4]

   (b) Create a new variable by combining the medium and large size rivers in one category. Is there a significant difference in mean chem2 value for rivers of small and medium/large sizes? Carry out an appropriate test. Include the appropriate hypotheses, test statistic value, p-value, and conclusion. [6]

2. Let X be a Bernoulli random variable with probability of success p (proportion), where X is 1 or 0 if the outcome is success or failure, respectively.

   (a) Construct a $(1-\alpha)100\%$ CI for p for large n (Analytically). [5]

   (b) Estimate the coverage probability of the standard 95% CI for proportion using Monte Carlo simulation. Use $n = 25, 40, 50$, and 120, and compare the results. [15]

3. Suppose $X_1, \cdots, _n$ follows an exponential distribution with scale parameter θ. We want to test $H_0 = \frac{1}{2}$ versus $H_1 \neq \frac{1}{2}$. To evaluate the sensitivity of non-normality find the Type I Error of the test for $n = 10, 30, 180, 200$ using Monte Carlo simulation. [20]

   Hints: It is a mean test. For each n, check how many times (from B iterations) you reject H_0 for a true H_0.