Last Name: _____ First Name: _____ Student ID: _____

AIDI 1000: AI Algorithms and Mathematics – Assignment - 1

Due Date : Feb 07, 2022, 11:59 PM

1. Simulate the Central Limit Theorem in any programming language. “The Central Limit Theorem states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger — no matter what the shape of the population distribution". (25 points)

2. Construct the binomial distribution for the total number of heads in four flips of a balanced coin. Define the PMF(Probability Mass Function) of the following distribution.(15 points)

3. Suppose that 40% of the voters in a city are in favor of a ban of smoking in public buildings. Suppose 5 voters are to be randomly sampled. Find the probability that (10 points):

   – 2 favor the ban.

   – less than 4 favor the ban.

   – at least 1 favor the ban.

4. Most graduate schools of business require applicants for admission to take the SAT examination. Scores on the SAT are roughly normally distributed with a mean of 530 and a standard deviation of 110. What is the probability of an individual scoring above 500 on the SAT? (15 points)

5. The Edwards's Theater chain has studied its movie customers to determine how much money they spend on concessions. The study revealed that the spending distribution is approximately normally distributed with a mean of 4.11 dollar and a standard deviation of 1.37 dollar. What percentage of customers will spend less than 3.00 dollar on concessions?(10 points)

6. A data scientist is testing a new model. She choose train and test sets at random from a large population of training data. She randomly choose 8 fold validation to get the accuracy for decision tree model, and choose 5 fold cross validation to get the accuracy for Logistic regression. The data are below: (25 points)

   – Decision Trees: 93,94,89,88,78,89,76,98

   – Logistic Regression: 78,90,89,76,89

   1. Are the two populations paired or independent? Explain your answer.
   2. Graph the data as you see fit. Why did you choose the graph(s) that you did and what does it (do they) tell you?
   3. Choose a test appropriate for the hypothesis above, and justify your choice based on your answers to parts (a) and (b). Then perform the test by computing a p-value, and making a reject or not reject decision. Do use python or any programming language for this, and show your work. Finally, state your conclusion in the context of the problem.