Homework 2

Statistical Inference

1- Two fair dice are rolled. A is defined as the event corresponding to "sum of two dice equals 4", B is the event corresponding to "sum of two dice equals 6", and C is the event corresponding to "at least one of the dice shows a 2".

   a. Calculate $P(A|C)$. Are A and C independent?

   b. Calculate $P(B|C)$. Are B and C independent?

2- Fifty-five percent of the students at a certain college are females. Eight percent of the students in this college are majoring in computer science. Three percent of the students are women majoring in computer science. If a student is selected at random, find the conditional probability that:

   a. The student is female given that the student is majoring in computer science.

   b. This student is majoring in computer science given that the student is female.

3- Many local polls of public opinion use samples of size 400 to 800. Consider a poll of 400 adults in Richmond that asks the question "Do you approve of President George W. Bush's response to the World Trade Center terrorists attacks in September 2001?" Suppose we know that President Bush's approval rating on this issue nationally is 92% a week after the incident.

   a. What is the distribution of random variable X that model number of people who answer positively to the above question? Explain.

   b. Calculate the probability that at most 358 of the 400 adults in the Richmond poll answer "Yes" to this question.

   c. Find the expected number of people in the sample who indicate approval. Find the standard deviation of X.

   d. Use a normal approximation to answer the question in (b), and compare the results of the exact value and the normal approximation. Is the normal approximation satisfactory?

4- You enter a chess tournament where your probability of winning a game is 0.3 against half the players (call them type 1) 0.4 against a quarter of the players (type 2) and 0.5 against the remaining quarter of the players (type 3). You play a game against a randomly chosen opponent. What is the probability of winning?

5- You write a software program over and over, and each time there is probability p that it works correctly, independent of previous attempts. What is the mean and variance of X, the number of tries until the program works correctly?

6- The time Negar takes to cycle from home to school is normally distributed with a mean of 40 and a variance of 7 (minutes). If Negar wants to have 95 percent confidence that she can attend her class at 1 pm, what is the latest time she can leave the house?

7- (R, ggplot) In this question, you are going to use the "Heart" dataset. This dataset consists of some details about 900 people. Note that you must use the ggplot2 library to draw the diagrams.

   a. Plot the histogram for **age** with an appropriate bin size. Then overlay that with the density curve. You also have to mark the 2.5% and the 97.5% percentiles on the diagram.

   b. Draw the QQ-plot of **thalch** (maximum heart rate achieved) for each **gender**. Can we sensibly assume that this variable is coming from a Gaussian distribution?

   c. Sort the categories in **origin** (place of study) by their frequencies, then draw a horizontal bar plot to show the result.

   d. Draw the separate boxplots of the **trestbps** (resting blood pressure) variable for each **cp** (chest pain type).

8- (R) In this question, we will learn more about Monte Carlo simulation and understand how it can help us to model and analyze complex systems, where analytical or closed-form solutions are difficult. We are going to estimate the value of π using Monte Carlo simulation. The idea is to randomly generate points within a square and determine how many fall in a circle inscribed by the square. By comparing the points inside the circle to the total number of points you can approximate π. Repeat your simulation with different number of points and analyze their results.