DASC 500 Homework 2

1. A box contains three marbles: one red, one green, and one blue.

   Consider an experiment where you draw a marble from the box, observe its color, and replace it on two successive attempts:

   - First Draw:
     - Randomly take one of the three marbles from the box
     - Note its color (C_1)
     - **Replace it in the box**
   - Second Draw:
     - Randomly take one of the three marbles from the box
     - Note its color (C_2)
     - **Replace it in the box**

   The result of the experiment is an ordered pair (C_1, C_2), where C_1 and C_2 can take on values of red, green, or blue.

   a. What is the sample space for this experiment?
   b. What is the probability of each point in this sample space?

   Assume now for the **same** experiment described above that marble selection order doesn't matter. That is, you consider drawing (R, B) to be equivalent to drawing (B, R), since both outcomes result in drawing one red marble and one blue marble.

   c. What is the sample space for the experiment when order does not matter?
   d. What is the probability of each point in the sample space when order does not matter?

2. Suppose cards numbered 1 through 10 are placed in a hat. After mixing up the cards, you draw a single card from the hat.

   If the number on the drawn card is at least 5, what is the conditional probability that the number on the drawn card is ten. (Hint: Let E denote the event that the number of the drawn card is 10 and let F be the event that it is at least 5.)

3. List the generic tasks included in the CRISP-DM Data Preparation Phase.

4. The Northwest regional manager of Logan Outdoor Equipment Company conducted a study to determine how her store managers allocate their time. The study lasted three weeks and documented the percentage of time spent on each task in the associated Excel file.

   a. Create a stacked bar chart of manager tasks based on store locations, with the locations along the vertical axis. Format the chart to best display these data to include axis labels, chart title, etc.

   b. Create a clustered-bar chart of manager tasks based on store locations, with the locations along the vertical axis. Format the chart to best display these data to include axis labels, chart title, etc.

Continued on next page

c. Create a single bar chart for each location. Each chart should show the percentage of time spent on tasks for that location. Format the bar charts to best display these data to include axis labels, chart title, etc.

d. Create a chart of your choosing that best provides insight into this data. Explain why you propose visualizing the data in this way.

e. What can we infer about both the similarities and the differences in the way store managers allocate their time at different locations?

Extra Credit Problem (worth 20 %)

Develop and execute a simulation of the marble-draw scenarios described in Question 1.

- Present the code used to simulate the scenario.
- Provide the results of the simulation and the estimated sample space probabilities.
- State whether the simulation results match your answers to both parts of Question 1 and explain why or why not.