Ec 5043 Prof. Le Wang
Managerial Economics II

Homework #1

Instruction: Do all the following empirical exercises using R. Turn in your answer with tables and graphs, if any, (along with your program and output files appended at the end of document). Refer to the output file whenever appropriate when discussing your results.

Note that for all simulation exercises, set the seed number to 123456 to ensure the reproducibility of your results.

Question 1. [Joint Distribution and Visualization]

Use the data titanic.csv on Github course page for this question. This dataset is frequently used in the machine learning literature as practice questions, containing all the information on Titanic passengers.

You can check out the following website for many hands-on examples using this dataset.
Machine Learning Basics with Application to Prediction of Survival on the Titanic

Below is a description of the dataset.

1. **Survived** Survival (0 = No; 1 = Yes)
2. **pclass** Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
3. **name** Name
4. **sex** Sex
5. **age** Age
6. **sibsp** Number of Siblings/Spouses Aboard
7. **parch** Number of Parents/Children Aboard
8. **ticket** Ticket Number
9. **fare** Passenger Fare
10. **cabin** Cabin
11. **embarked** Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

Answer the following questions:

1. Using R, report the joint distribution (probability mass function) for survival status and passenger class. Interpret the results. Do these results make sense? **This question is corresponding to the definition of PMF in your slides.**
2. Using R, calculate marginal distributions for both survival status and passenger class as well. **This question is corresponding to "Use 3: Marginal Distribution" using the PMF in your slides.**
3. Plot a Mosaic plot to visualize the relationship between survival status and passenger class. **This question is corresponding to Visualization in R (Mosaic Plot) in your slides.**

4. What is the probability of survival for passengers in the first and second classes? **This question is corresponding to "How can we further use the joint p.mf.?" and "Use 2: Cumulative Distribubtion Function"**
5. Conduct the test of independence between survival status and passenger class in R.
   (a) Write down the **Expected Table**. **This question is corresponding to the definition of Expected Table and the example for political affiliation and opinion on tax reform.**
   (b) Manually calculate the test statistic using your **Expected Table**. **This question is corresponding to the example right under the test statistic in your slides.**
   (c) Report and intepret your results. **This question is corresponding to the "conclusion" of test of independence in your slides.**
   (d) Propose a way to visualize independence, without using statistical tests, such as those suggested by Heather and Richard in class. **This question is corresponding to "How to visualize independence" in your slides.**

Question 2. [Joint Distribution and Portfolio Choice]

Consider a young investor who owns some stock shares in two companies. Specifically, she owns 100 shares of the stock A that will be worth X and 200 shares of the stock B that will be worth Y. In other words, the total value of her portfolio W is given by the following formula:

$$W = 100 \times X + 200 \times Y$$

To understand the potential returns to her portfolio, we need to have the information on the joint distribution of the stock prices for these two companies, which can be esitmated from historical data and given in the following table. To simplify our analysis, we assume that the prices for each stock can take on only three different values.

X/Y	30	32	34
70	0.1	0.1	0
75	0.2	0.2	0
80	0	0.2	0.1

Table 1: Joint Probability Distribution of Two Stocks

1. Verify if the table actually represents a joint probability distribution table. Explain your results. **This question is corresponding to the properties of a joint probability mass function in your slides.**
2. What is the probability of having returns to Stock A: $X \leq 75$? **This question is corresponding to "Use 3: Marginal Distribution" in your slides.**
3. She plans to sell these shares next year for a down payment on a condominium, which is about $13,000$ dollars. What is the probability that her portfolio is actually worth $13,000$? **This question is corresponding to the definition of joint probability mass function in your slides.**

Question 3. [Emergency Treatment of Septic Shock]

From Paul R. Rosenbaum Observation and Experiment: Septic shock occurs when a widespread infection leads to very low blood pressure. It is often lethal, particularly in young children and the elderly. Each year in the United States, there are more than $750,000$ cases of spetic shock. The initial treatment for septic shock typically occurs in a hospital emergency room, and 20% patients may die.

	In-hospital 60-day Mortality	
Treatment Group	In-hosipital Death	Other
Aggressive	92	347
Less Aggressive	81	365

Table 2: In-Hospital Mortality Outcomes

What is the best way to treat septic shock? Some researchers proposed an agressive, six-hour protocol for treatment of septic shock. In 2014, the New England Journal of Medicine reported the results of a randomized clinical trial comparing the aggressive protocol with a less-aggressive, six hour protocol of standard therapy.

1. Use R to test whether or not the treatment is actually related to (or dependent of) in-hospital 60-day mortality rate. Later we will use data like this to evaluate whether or not the treatment is effective or not, and how large the effect is.