DUE: October 10th, 2018 at 11:59 PM

QMM 1001 - Statistics for Data Analytics Lab 3 /30

Create and submit an R script which, when run, will print the answers to the following questions. Your R script must include a title with your name and student number and comments for each question number and letter.

1. **(11 marks)** The data set "Car.Discounts.csv" shows car discounts based on a person's age, income, and gender. Read this dataset into R and use it to answer the following questions.
    a. Find the mean of the discount variable. (1 mark)
    b. Find the range of the discount variable. (1 mark)
    c. Find the standard deviation of the discount variable. (1 mark)
    d. Find the quartiles of the discount variable. (1 mark)
    e. Find the mean discount for males (0 = male). (2 marks)
    f. Find the mean discount for females (1 = female). (2 marks)
    g. Create a boxplot that shows the discount for each gender in the same plot. Create an appropriate title and colour the box for females pink and the box for males blue. What can you infer from this boxplot? (3 marks)

2. **(11 marks)** The data set "House.Price.csv" shows house price and the amount of living space of the house in square feet. Read this data set into R and use it to answer the following questions.
    a. Create a histogram of the house price variable with 20 bins. Add an appropriate title and x and y axis labels. Change the color of the bars to any color other than white and the color of the borders to any color other than black. (4 marks)
    b. Describe the shape of this distribution (symmetric, left skewed, right skewed). Find the mean and median for the house price variable. How do these values confirm the shape of the distribution? (4 marks)

c. Transform the house price variable using a log transformation. Create a histogram for the log of house prices. What do you notice about the shape of the distribution? (3 marks)

3. **(8 marks)** The data set "Workforce.csv" shows the year, annual average workforce participation (the percentage of the population 16 years and older that is employed or unemployed), male workforce participation (the annual average for males only), and female workforce participation (the annual average for females only). Read this data into R and use it to answer the following questions.

   a. Create a time series object for the annual average workforce participation, male workforce participation, and female workforce participation using the ts() function. This time series object should start in year 1948 and end in year 2015. (3 marks)

   b. Plot the time series data. Add an appropriate title. (2 marks)

   c. Briefly describe the trends in average workforce participation, male workforce participation, and female workforce participation over time. (3 marks)

Save your R Script as: Last Name, First Name LAB 3

Upload your R Script to the "R Assignment – Lab 3" drop box on Moodle before October 10th at 11:59 PM.