DSO 499: Business Analysis using R

Final Exam Sample Test

Exam length: 120 min

Instructions:

Download data before you start the test. There are 3 data sets for the exam, please download all of them before you begin the test.

The final exam is comprehensive. It consists of 3 problems. You'll have to use R for all problems.

The test must be completed in one sitting and will be timed. You will have 120 minutes to complete the exam.

Save your script often. You will be asked to submit the script.

Problem 1: Use R to answer the following questions. Dataset for this problem can be found in Problem1.xlsx file

1. Use R to calculate the following summary statistics for the variable "Salary":
    a) Mean
    b) Median
    c) 1st Quartile
    d) 3rd Quartile
    e) Standard deviation
    f) IQR
    g) Minimum
    h) Maximum
    i) Range

2. How many employees work in the Marketing Department?
3. How many employees have a salary of at least $75,000?
4. What is the median salary in the Marketing Department?
5. Create a histogram of employee salaries. Submit the PDF of the histogram here.

Problem 2: Use R to answer the following questions. Data for this problem can be found in *Problem2.xlsx* file.

The Marlins General Manager is disgruntled because two desirable rookies accepted offers from the Yankees instead of the Marlins. He believes that Yankee salaries must be noticeably higher—otherwise, the best players would join the Marlins organization. If the typical Yankee is better compensated, the General Manager is planning to chat with the Owners about sweetening the Marlins' offers.

Perform the appropriate hypothesis test to help Marlins' manager decide. Answer the following questions:

1. You decided to check whether there is evidence that Yankee salaries are higher. What is the name of the appropriate hypothesis test?
2. What is the value of the test statistic?
3. What is the p-value?
4. Is the test significant at significance level of 0.05?
5. Is a typical Yankee better compensated?

Problem 3: The data in *Problem3.xlsx* lists information about 400 customer transactions. It contains information on the transaction date, day of the week, time of the day, region, payment type, gender of the customer, and total cost. Answer the questions below using dplyr, ggplot2 packages, and pipes approach.

1. What is the total amount spent by female customers?
2. Report the R code you used to calculate the total amount spent by female customers, make sure you used dplyr package and pipes.
3. What is the average amount spent by female customers?
4. Report the R code you used to calculate the average amount spent by female customers, make sure you used dplyr package and pipes.
5. Create a table from this data set that shows the number of transactions done by **gender** and **time of day**. Report the R code you use for that (use pipes and dplyr package).
6. Create a boxplot of the amount spent split by gender. Use ggplot2 package. Submit the PDF file of the figure.
7. Submit the R code that produces a boxplot of the amount spent split by gender from previous question.