CS130 Assignment #3

Date Assigned: Monday, October 9, 2017
Date Due: Monday, October 16, 2017
Points: 50

Create a Word document **PUNuetIDAnswers.doc** using your PUNetID that will contain answers to each of the following questions in order when applicable. Your answer document and any other files created are to be placed in a folder called **PUNetIDSolution3** once again using your PUNetID. The entire folder **PUNetIDSolution3** is to be dropped in the CS130Drop folder on grace by 2:15pm on the day the assignment is due. **There is no late grace period for this last assignment.**

Problem #1

Create Worksheet #5 - named **S&P500**

Using the graph at https://finance.yahoo.com/quote/%5EGSPC?p=^GSPC (select Max) find the following:

Date	S&P 500 Value
Jan 1, 1950	
Jan 1, 1960	
Jan 1, 1970	
Jan 1, 1980	
Jan 1, 1990	

1) Enter the data into worksheet #5 of your existing file of worksheets so that x represents the Years since Jan 1, 1950. Column headings are to be **Years Since 1950** (e.g. enter x = 6 for 1956) and **Price Per Share**. Make your worksheet looks professional and display the S&P 500 values to in dollars to 2 decimal places.

2) Perform an exponential regression.

3) Paste a copy of the Scatterplot with the regression equation and R^2 value in your Word document under a heading **Problem #1 - Question 1**.

4) Would this correlation be considered a "strong" correlation based on our discussions in class? Why or why not? Answer this question in your Word document under the heading **Problem #1 - Question 2**.

5) What is the predicted per share value of the S&P500 on Jan 1, 2010? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 3**. Show your work for full credit.

6) What is the actual per share value of the S&P500 on Jan 1, 2010? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 4.**

7) Add in the correct row data for Jan 1, 2000 and Jan 1, 2010 and update all Chart information. Paste your updated chart under the heading **Problem #1 - Question 5**.

8) What is the predicted per share value of the S&P500 on Jan 1, 2017? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 6**. Show your work for full credit.

9) What is the actual per share value of the S&P500 on Jan 1, 2017? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 7.**

Note: All answers to this question are to fit on page one of the Word document and answered in order.

Problem #2

Using the sample data set entitled "IceCreamSales.xlsx" found in the CS130Public folder and R, answer each of the following questions in the Word document PUNetIDAnswers.doc as specified. When asked for, place a graph into your document with the appropriate explanation.

0) Create a directory **c:\r-data.** Copy IceCreamSales.xlsx into the new directory.

1) For each variable in IceCreamSales.xlsx, identify a) variable name, b) qualitative or quantitative, c) discrete, continuous, nominal, or ordinal. Answer this question under a heading of **Problem #2 - Question 1**.

2) Write an R script called IceCreamSales.R that does the following:

a) Changes the location of your working directory to **c:/r-data**.

b) Installs the Readxl package and then activates the package

c) Imports the IceCreamSales.xlsx into a variable called icecreamsales

d) Computes and shows on the screen the mean and standard deviation of the variables FarenheitTemperature and IceCreamSales.

e) Create a Scatterplot of FarenheitTemperature and IceCreamSales. Yes, you will have to look this up. It's very straight forward given what we've learned in class. Make sure your plot is properly titled (i.e. plot title, proper x-axis label, and proper y-axis label). Paste in your answer to this question under a heading of **Problem #2 - Question 2**.

Note: I will run your R script from the directory c:\r-data of a comparable PC to Scott 204. If the script does not work correctly, you will lose significant points. Mac users make sure to test your script on a PC in Scott 204.

3) Which variable is independent? Why? Give a detailed explanation for full credit. Answer this question under a heading of **Problem #2 - Question 3**.

Note: All answers to this question are to fit on page two of the Word document and answered in order.

Problem #3

A pretty comprehensive list of caffeinated drinks can be found in the Excel file EnergyDrinks.xlsx in the CS130Public folder on grace. Using the sample dataset and R, answer each of the following questions in your Word document PUNetIDAnswers.doc as specified. When asked for, place a graph into your document with the appropriate explanation.

1) Copy EnergyDrinks.xlsx into c:\r-data.

2) Write an R script called EnergyDrinks.R that does the following:

a) Changes the location of your working directory to **c:/r-data**.

b) Installs the readxl package and then activates the package.

c) Imports the EnergyDrinks.xlsx file into a variable called energyDrinksFrame

d) Finds the mean and standard deviation for the column of data representing mg/floz. Paste the table of values created by R in your Word document under the heading **Assignment #3 - Question 1**.

3) Create a bar graph that shows the number of Coffee, Energy Drinks, Energy Shots, Other, Soda, and Tea totals in the entire dataset. Display the bars in different colors and in alphabetical order. Make sure your bar graph has a proper title and appropriate axis names. Paste your graph under the heading **Assignment #3 – Question 2**.

4) Question: Assuming a normal distribution, is the milligrams per ounce the same for energy shots and energy drinks. Under the heading **Assignment #3 – Question 3**, answer each of the following questions in order in your Word document:

a) What hypothesis test will you use to answer this question? Explain in detail why you selected the test you did.

b) State the Null Hypothesis.

c) Perform the correct hypothesis test.

d) Paste in the results produced by R for the hypothesis test stated in b).

e) Do you accept or reject the Null Hypothesis? Why?

f) State your conclusion.

For questions a) through f) above, first type **4 a)** and then place your answer for a), and so on.

5) Question: You are to determine whether the mean milligrams per fluid ounce of only the sodas is different than 2.75 in a statistically significant way.

Under the heading **Assignment #3 – Question 4**, answer each of the following questions in order in your Word document:

a) What hypothesis test will you use to answer this question? Explain in detail why you selected the test you did.

b) State the Null Hypothesis.

c) You will need to grab only the soda observations from the Excel file and place this data in an excel file called **Soda.xlsx**. Then add to the already written script the logic to perform the correct test.

d) Paste in the results produced by R for the hypothesis test specified in b).

e) Do you accept or reject the Null Hypothesis? Why?

f) State your conclusion.

For questions a) through f) above, first type **5 a)** and then place your answer for a), and so on.

Note: All answers to this question are to fit on pages three and four of the Word document and answered in order.

How to Submit and Grading Policies

A copy of your single **<u>folder</u>** (PUNetIDSolution3) with all of the files (properly named) is to be placed in the CS130Drop folder by **2:15pm** on the due date to be considered on time.

Here are the files I'm expecting in the folder PUNetIDSoltuions3

1. PUNetIDAnswers.doc
2. PUNetIDExcel.xlsx with the addition of worksheet #5
3. IceCreamSales.R
4. EnergyDrinks.R
5. Soda.xlsx

Further, you are to turn in a printed copy of PUNetIDAnswers.doc. Page 1 of the document is to contain ALL answers to Problem #1. Page 2 is to contain ALL answers to Problem #2 and page 3 & 4 is to contain ALL answers to Problem #3. Do not use less than a 10-pt font size. Print this document in color (double-sided and stapled in the upper left corner) and place on the instructor's desk no later than 2:15pm on the day the assignment is due.

Grading will be based on:

• 	Correctness of your results

• 	Completeness of your results

• 	Professional look of the worksheets, charts, and document as described above and discussed in class

• 	Ability to perform a what-if analysis in any worksheet by changing any of the user input data with accurate results computed and displayed in the worksheet.