BTMA 531 Assignment 1:
R Basics and Regression

Due January 29, 2020 by noon on D2L

Instructions

- You should create a single R script or an R Notebook called {firstname}_{lastname}_Asgn1.r, which has the required code for all parts of the assignment.
- Make sure to use commenting (#) so that your R file is readable by someone else. Yo do not need to comment on what you are trying to do in each line, but it should be clear where the answer for each question is.
- Make sure that your R script is executable from top to bottom on another computer. A good approach is to test it before submission on a lab computer. Make sure all requirements of questions are done using R (e.g. do not use calculator, excel, ... to calculate things).
- The assignment submission should be done through D2L dropbox. Upload the R file to the assigned dropbox folder in D2L.
- The purpose of the assignments is to help you learn through practice. I recommend working on assignments in groups. You may want to use R help or search online for answers. However, note that this is an individual assignment. The work you submit should be 100% yours. Do not copy, share, or ask for files, chunks of code, or answers. Refer to the course outline for some examples of what to do and what not to do, and to Code of Student Conduct for more information on cheating and plagiarism. If you are note sure about a behavior, please ask.

Questions

1 Create a sequence of numbers starting from 1 and going up to 10 by increments of 0.5. Assign this sequence to an object called "object1". What is the answer to following command? [5]

```
object1[14]
```

2 List out all objects in the environment. Remove either "object1" or "object2" from the environment. [5]

3 What do you need to do to use the funtions in "dplyr" package using code only? [5]

4 Load the package "MASS" and open the "cabbages" dataset. How many variables does the dataset have? How many of these variables are qualitative/categorical and how many are quantitative (answer programmatically)? [10]

5 In the dataset "cabbages", select the data for which the cultivar (Cult) is "c39". [5]

6 In the dataset "cabbages", calculate the average cabbage head weight (HeadWt) of cabbages for which cultivar is "c39". [5]

7 From dataset "cabbages":[20]

a) Graph a scatter plot of cabbage head weights and their vitamin C (VitC) content.

b) Properly name the x and y axes.

c) Add a linear estimate of the relationship to the graph.

d) Find the correlation between these two variables.

8 From dataset "cabbages", create frequency tables for "Cult" and "Date" variables. How many entries are there for each Cult type and each Date, respectively? [5]

9 From dataset "cabbages":[10]

a) Create two plots: one bar plot of the "Cult" variable, and one pie plot of the "Date" variable, and put them side by side (both plots in one figure).

b) create a histogram of the variable "VitC" (one single graph filled by the figure).

10 [20]

a) Import the attached "ad.csv" file to R as an object called "data", once using read.csv and once using fread. Add libraries if needed.

b) Create a simple linear regression model for sales using TV advertising variable. Does TV advertising impact sales? Predict the sales for when TV advertising is 75.

c) Create a multiple linear regression model for sales using TV, radio, and newspaper advertising variables.

d) Draw the normality plot of residuals for the multiple linear regression. What does it say?

11 From dataset "cabbages": [10]

a) Create a multiple linear regression model (with categorical variables) that uses "Cult", "Date", and "HeadWt" to predict "VitC".
b) Predict vitamin C of a cabbage with HeadWt of 2, Date of d16, and Cult of c52.