Midterm

STAT 471/571/701 Modern Data Mining

03/25/2019

Contents

Name your submission using the scheme:

`LastName_FirstName_Midterm`

For example: `Zhao_Linda_Midterm` .rmd, **.pdf**, **.html** or **.docx**.

Instruction: This exam requires you to use R. It is completely open book/notes. Write your answers using .rmd format and knitr it into one of the html/pdf/docx format. Show your codes, plots or R-output when needed. If you have trouble formatting the plots, don't worry about it. We are not looking for pretty solutions, rather to see if you are able to make sense out of data using R.

Data and the midterm exam for Midterm can be found at:

- Data needed: /canvas/Files/Midterm/data_exam.csv

- Midterm with different formats:

    - /canvas/Files/Midterm/midterm_spr19.Rmd

    - /canvas/Files/Midterm/midterm_spr19.pdf

    - /canvas/Files/Midterm/midterm_spr19.docx

- Help: As always skip any part you have trouble with and you may come back to finish it if you have time. Ask one of us for help if you are stuck somewhere.

- Electronic submission: In the 'Assignments' section, go to the 'Midterm' assignment and upload your completed files. It takes multiple files. The folder will be closed at 8:10 pm.

- If you have trouble to upload your files, email them to lzhao@wharton.upenn.edu and arunku@wharton.upenn.edu

Introduction

```
### WE USE tips.csv FOR YOU TO GET FAMILIAR WITH THE PROCESS
### IT IS NOT THE DATA WE WILL USE IN THE EXAM
data.exam <- read.csv("tips.csv",
                      header = T, na.strings = c("?", ""))
```

Question 1

a)

i)
```
# code block if necessary
```

ii)

iii)

b)

i)

ii)

iii)

Question 2

BLAH. . .

. . . .

Declaration

By submitting this document you certify that you have complied with the University of Pennsylvania's Code of Academic Integrity, to the best of your knowledge. You further certify that you have taken this exam under its sanctioned conditions, i.e. solely within the set exam room and within the time allotted.