PH251D Fall 2018 - Project 1

FirstName MI LastName

10/29/2018

Create a project folder called project1 on your computer. You will put all your Project 1 files in this folder.

Go to my GitHub site at <https://github.com/taragonmd/data>.

Go into the project1 folder.

Download this Rmarkdown template (PH251D2018\_LastName\_Project1.Rmd) and edit. Use R Markdown to demonstrate the following skills:

## 1. Using the source function

Download the problem1.R file and save to the project1 folder. Run the program file (problem1.r) using the ‘source’ command. Show the R code chunk and results below.

source('problem1.R')

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 4

## 2. Read an ASCII data set

The Evans data set (evans.txt) is here: <https://github.com/taragonmd/data>.

Alternatively, here is the raw Evans data set: <https://raw.githubusercontent.com/taragonmd/data/master/evans.txt>.

Demonstrate reading the Evans data file (evans.txt) to create a data frame, and use the str function to explore the structure of the data set. Show the R code chunk and results below.

edat <- read.table('https://raw.githubusercontent.com/taragonmd/data/master/evans.txt',  
sep='', header=TRUE)  
str(edat)

## 'data.frame': 609 obs. of 12 variables:  
## $ id : int 21 31 51 71 74 91 111 131 141 191 ...  
## $ chd: int 0 0 1 0 0 0 1 0 0 0 ...  
## $ cat: int 0 0 1 1 0 0 0 0 0 0 ...  
## $ age: int 56 43 56 64 49 46 52 63 42 55 ...  
## $ chl: int 270 159 201 179 243 252 179 217 176 250 ...  
## $ smk: int 0 1 1 1 1 1 1 0 1 0 ...  
## $ ecg: int 0 0 1 0 0 0 1 0 0 1 ...  
## $ dbp: int 80 74 112 100 82 88 80 92 76 114 ...  
## $ sbp: int 138 128 164 200 145 142 128 135 114 182 ...  
## $ hpt: int 0 0 1 1 0 0 0 0 0 1 ...  
## $ ch : int 0 0 1 1 0 0 0 0 0 0 ...  
## $ cc : int 0 0 201 179 0 0 0 0 0 0 ...

## 3. Discretizing a continuous variable into a categorical variable

Total cholesterol levels less than 200 milligrams per deciliter (mg/dL) are considered desirable (**normal**) for adults. A reading between 200 and 239 mg/dL is considered **borderline high** and a reading of 240 mg/dL and above is considered **high**.[[1]](#footnote-26)

The Evan data dictionary is in Appendix D of the PHDSwR book. Convert total cholesterol variable (chl) into a categorical variable (factor) with the three levels described above.

cat('Edit this R chunk')

## Edit this R chunk

## 4. Working with dates and times

President John F. Kennedy was assassinated on “November 22, 1963”. Convert this character string into a R date object. Show how to use R to display (a) the Julian date; (b) the day of the week, and (c) the week of the year.

cat('Edit this R chunk')

## Edit this R chunk

## 5. Simple two-way analysis

Create a simple 2x2 table of smoking (smk) and coronary heart disease (chd). Use the fisher.test on this 2x2 table and describe your findings.

cat('Edit this R chunk')

## Edit this R chunk

## 6. Write your own function

Now, write a function to calculate the odds ratio of your 2x2 table above.

cat('Edit this R chunk')

## Edit this R chunk

## 7. Nested for loops

Write a nested for loops to create a mulitiplication table for the numbers 1 to 10.

cat('Edit this R chunk')

## Edit this R chunk

## 8. Create a simple graph

From the Evans data create a histogram of the total cholesterol (chl). Label with a title and axis labels. Output to a PNG file using the png function.

cat('Edit this R chunk')

## Edit this R chunk

## 9. Display PNG file in your Rmarkdown document

Using Rmarkdown syntax, display the PNG you created above.

cat('Edit this R chunk')

## Edit this R chunk

## 10. Using regular expressions

Here are the California counties: <https://github.com/taragonmd/data/blob/master/calcounty.txt>

Remove the “California” entry.

Use regular expressions to identify and display the County names that start with two or three letters followed by a space (e.g., "San ").

cat('Edit this R chunk')

## Edit this R chunk

1. Source: <https://www.medicalnewstoday.com/articles/315900.php> [↑](#footnote-ref-26)