Assignment 3
Logicals and simple data frames manipulation
Wednesday Feb. 5, 2014
Due date: Fri. Feb. 14, 2014

In this assignment, we work with logical expressions and simple dataframes. We work on extracting subsets of the data, and on putting the commands used inside a function, that is inside a script. There should be one script for each function in this assignment. These scripts should have names such as "question1.r", etc., and collected into a single archived file via tar or zip or gzip and emailed to the TA by the due date.

Question 1: 25 pts

(1) Generate 10000 random even integers in the range 2 to 1024. Extract a vector with the random numbers less than 500 and calculate the mean of this vector.

(2) From the same vector of 10000 random even integer above, extract all integers that are either less than 380 or greater than 882. What is the mean value of the resulting extracted vector?

Question 2: 25 pts
Using the `data()` function, read the dataset named `USJudgeRatings` and provide the following information: (In each case, write the command that allowed you to provide the answer to the question.)

(a) What is the type (or class) of the datataset

(b) What is the structure of the dataset? How many columns? How many rows in each column?

(c) What is the type of each column? What are the column headers?

(d) Compute the mean and standard deviation of judge integrity for judges with diligence above the mean. Repeat for judges with diligence below the mean. Conclude whether or not diligence affects integrity or not. Explain your answer (a simple yes or no is not sufficient.) Use the help system to learn more about the dataset. (Note: I do not know what to expect).

(e) Repeat (d) to find out whether "Preparation for Trial", and "Demeanor" has any relation (i.e., correlation) with judge integrity or not.

Question 3: 25 pts
Use the data from "USAccDeaths". What type of data is it? What is its structure? Using help or google, write some text explaining the type of structure of this data.

Question 4: 25 pts
Read the data.set named esoph that contains the following information from a case-control study of esophageal cancer in Ile-et-Vilaine, France (use ?esoph to get this information).

(1) What is the structure of this dataset? What are the types of each column?

(2) Compute the mean number of controls for each age group: 25-34, 35-44, and 45-54.

(3) Compute the mean number of controls for the combination of age groups: 65-74 and 75+.

(4) Compute the total number of cases for all patients age 55 and above.

(5) Create a new dataframe composed of only the columns the correspond to the Age Group and Tobacco consumption. Store the dataframe in a variable of your choice.

(For this assignment, youll be using most of the following functions: (mean, sd, ?, data.frame, source, str, class, sample).)

Provide a 250 word summary of this assignment. Also provide a single script for the four questions. Inside the script, use the command:

```
print("Question 1")
```

for each question, followed by the commands that must be run to generate the answers to the required question and the required plots.