Homework 1 - Statistics 20

Due April 8, 2016 before 11:59 PM via upload to CCLE

Installing R & RStudio

Problem 1. Please install the most recent versions of R (3.2.1) & RStudio on your computer. As proof of completion, issue this commnd if you use MacOS:

```
system("who", intern = TRUE)
```

```
[1] "Nancy    console  Mar 15 09:21 "
```

but if you use Windows:

```
system("whoami", intern = TRUE)
```

```
[1] "Nancy"
```

Swirl Package

Problem 2. In this problem, we'll be using the swirl software package for R in order to practice some key concepts. The swirl package turns the R console into an interactive learning environment.

- Install swirl using: **install.packages("swirl")**
- Load Swirl using: **library(swirl)**
- Install the R Prograomming course using: **install_from_swirl("R Programming")**
- Start swirl and complete the lessons using: **swirl()**
- Do the following lessons:

A. Do *1: R Programming Basic Building Blocks* lesson. Explain the *recycling* rule with an example.

When given two vectors of the same length, R simply performs the specified arithmetic operation (+, -, *, etc.) element-by-element. If the vectors are of different lengths, R 'recycles' the shorter vector until it is the same length as the longer vector. e.g.

```
z<-c(1.1,9,3.14)
z*2+100
```

```
[1] 102.20 118.00 106.28
```

When we do z * 2 + 100, z is a vector of length 3, but technically 2 and 100 are each vectors of length 1. So R is 'recycling' the 2 to make a vector of 2s and the 100 to make a vector of 100s.

B. Do *3: Sequences of Numbers* lesson. What command was suggested for generating a vector to contain 10 zeros, then 10 ones, then 10 twos?

```r
rep(c(0,1,2),each=10)
```

```
 [1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
```

C. Do *4: Vectors* lesson. What was the output for paste(my_name,collapse=" “)?

```r
my_char<-c("My","name","is")
paste(my_char, collapse = " ")
```

```
[1] "My name is"
```

D. Do *7: Matrices and Data Frames.* What was the output for colnames(my_data)?

```r
my_vector<-1:20
dim(my_vector) <- c(4, 5)
my_matrix<-my_vector
patients<-c("Bill", "Gina", "Kelly", "Sean")
cbind(patients,my_matrix)
```

```
     patients
[1,] "Bill"   "1" "5" "9"  "13" "17"
[2,] "Gina"   "2" "6" "10" "14" "18"
[3,] "Kelly"  "3" "7" "11" "15" "19"
[4,] "Sean"   "4" "8" "12" "16" "20"
```

```r
my_data <- data.frame(patients, my_matrix)
cnames<-c("patient", "age", "weight", "bp", "rating", "test")
colnames(my_data)<-cnames
colnames(my_data)
```

```
[1] "patient" "age"     "weight"  "bp"      "rating"  "test"
```

Character (String) and Logical Vectors

Problem 3. Suppose a student has written the following:

```r
a <- c("Jane", "Dave", "Ann")
b <- c(TRUE, FALSE, FALSE)
b+5
```

```
[1] 6 5 5
```

```r
d <- matrix(c(a,b), nrow=3)
d[,2]+5
```

```
Error in d[, 2] + 5: non-numeric argument to binary operator
```

Why does d[,2]+5 result in an error but b+3 did not? This answer only requires an explanation, no code.

When we calculate b+3, R automatically converts the logical value into a numeric, so we are acturally calculating c(1,0,0)+5. However, when we combine a and b into a matrix, R automatically converts the logical values into characters, so we are acturally calculating c(“TRUE”, “FALSE”, “FALSE”)+5, which is not allowed.

Matrices

Problem 4. For a matrix

```r
X <- matrix(56:100, ncol=9)
X
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]   56   61   66   71   76   81   86   91   96
[2,]   57   62   67   72   77   82   87   92   97
[3,]   58   63   68   73   78   83   88   93   98
[4,]   59   64   69   74   79   84   89   94   99
[5,]   60   65   70   75   80   85   90   95  100
```

A. Show how to display the values in the 2nd column only.

```r
X[,2]
```

```
[1] 61 62 63 64 65
```

B. Show how to display the values in 3rd and 5th rows only.

```r
X[3,]
```

```
[1] 58 63 68 73 78 83 88 93 98
```

```r
X[5,]
```

```
[1]  60  65  70  75  80  85  90  95 100
```

Data Frames

Problem 5. Data frames share some of the properties of matrices and of lists. Technically, they are considered lists (generic vectors) in the R language. They are used as the fundamental data structure by most of R's estimation functions.

A. Demonstrate that you are able to use the function data() and that you know what a data frame is by finding a built-in dataset that is NOT USArrests. Please identify the dataset. For example:

```r
data() #show all the built-in datatsets
data(BOD)
```

Then do the following:

B. Issue the class() and mode() functions for your data frame so that we can see you truly selected an appropriate dataset.

```r
class(BOD)
```

```
[1] "data.frame"
```

```
mode(BOD)
```

```
[1] "list"
```

C. How many observations and how many variables does your data frame have? Use any R function to reveal this information.

```
nrow(BOD) #number of observations
```

```
[1] 6
```

```
ncol(BOD) #number of variables
```

```
[1] 2
```

D. Apply the summary function to a column of your choice.

```
summary(BOD[,2])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   8.30   11.62   15.80   14.83   18.25   19.80 
```

E. Apply the head function to a column of your choice.

```
head(BOD[,2])
```

```
[1]  8.3 10.3 19.0 16.0 15.6 19.8
```