R Homework 6

November 6th, 2018

Fisher Ankney

Statistics 5193

Question 1.

Consider the social media data on Brightspace. Load the data into R and save it as a data frame called "your initials". It should have row labels 1, 2, . . . , 35 and appropriate column names.

```
library(readxl)
StudentData <- read_excel("/Users/fisher/Documents/data_science/r_stat_5193/data/StudentData.xlsx")
fra <- data.frame(StudentData)
row.names(fra) <- seq(1:35)
```

Question 2.

Observe the first person in the data set has 1 text sent and 1 text received. This person later told me that they didn't have a cell phone. Change this value to NA for both texting variables and print the first 3 rows of the data set to the console.

```
fra$TxtSent[1] <- NA
fra$TxtRec[1] <- NA

fra[1:3,]
```

```
##   Gender    Class HSClass TxtSent TxtRec Fbtime Pinterest Snapchat
## 1      M STAT2023       1      NA     NA     30         N        Y
## 2      M STAT2023      15      10     15     20         N        Y
## 3      M STAT2023      65     150    150     80         N        Y
##   Introvert
## 1         8
## 2         8
## 3         1
```

Question 3.

Add a new variable to the data set called "d.text" that computes TextSent-TextRec, put the new variable between TxtRec and Fbtime variables, and print the first 3 rows to the console.

```
d.text <- fra$TxtSent - fra$TxtRec
fra <- data.frame(fra[1:5], d.text, fra[6:9])

fra[1:3,]
```

```
##   Gender    Class HSClass TxtSent TxtRec d.text Fbtime Pinterest Snapchat
## 1      M STAT2023       1      NA     NA     NA     30         N        Y
## 2      M STAT2023      15      10     15     -5     20         N        Y
## 3      M STAT2023      65     150    150      0     80         N        Y
##   Introvert
## 1         8
## 2         8
## 3         1
```

Question 4.

Create a new variable called d.text.cat that is an ordered factor indicating when d.text is less than 0, 0, or greater than 0. The values of the factor should be "Neg","0","Pos" and NA. Also d.text.cat should be added to the data set. Print the data set to the console.

```
d.text.cat <- factor(sign(d.text), ordered = TRUE,
                     levels = c(-1, 0, 1),
                     labels = c("Neg", "0", "Pos"))

fra$d.text.cat <- d.text.cat

addNA(fra$d.text.cat)
```

```
head(fra)
```

```
##   Gender    Class HSClass TxtSent TxtRec d.text Fbtime Pinterest Snapchat
## 1      M STAT2023       1      NA     NA     NA     30         N        Y
## 2      M STAT2023      15      10     15     -5     20         N        Y
## 3      M STAT2023      65     150    150      0     80         N        Y
## 4      F STAT2023     123      18     28    -10     45         N        Y
## 5      F STAT2023     130      30     30      0     20         N        Y
## 6      F STAT2023     140     100     75     25     60         Y        Y
##   Introvert d.text.cat
## 1         8       <NA>
## 2         8        Neg
## 3         1          0
## 4         4        Neg
## 5         4          0
## 6         6        Pos
```

Question 5.

Order the data set by gender (ladies first) and d.text (ascending) within gender. Save the ordered data set as "Ordered" and print it to the console.

```
Ordered <- fra[order(fra$Gender,fra$d.text),]
head(Ordered)
```

```
##    Gender    Class HSClass TxtSent TxtRec d.text Fbtime Pinterest Snapchat
## 17      F STAT2023     760      20     40    -20     10         Y        Y
## 4       F STAT2023     123      18     28    -10     45         N        Y
## 26      F STAT5063      40      10     20    -10      5         N        Y
## 29      F STAT5063      50      11     15     -4     80         N        N
## 30      F STAT5063      50       2      3     -1     10         N        N
## 5       F STAT2023     130      30     30      0     20         N        Y
##    Introvert d.text.cat
## 17         3        Neg
## 4          4        Neg
## 26         5        Neg
## 29         3        Neg
## 30         4        Neg
## 5          4          0
```

Question 6.

Students 1 – 22, 25, and 30 had taken some undergraduate courses at OSU while the rest had not. Create a variable called U.OSU that takes on values Y for students 1 – 22, 25, and 30 and N otherwise. Use the rownames function or the merge function, or both, to merge the variable to the ordered data set in 5. Print the first 5 rows to the console.

```
U.OSU <- rep('Y', 35)
U.OSU[23:24] <- 'N'
U.OSU[26:29] <- 'N'
U.OSU[31:35] <- 'N'
U.OSU <- data.frame(U.OSU)

index <- as.numeric(row.names(Ordered)) # used row.names to index

Ordered <- cbind(Ordered, U.OSU[index,])
Ordered[1:5,]
```

```
##    Gender    Class HSClass TxtSent TxtRec d.text Fbtime Pinterest Snapchat
## 17      F STAT2023     760      20     40    -20     10         Y        Y
## 4       F STAT2023     123      18     28    -10     45         N        Y
## 26      F STAT5063      40      10     20    -10      5         N        Y
## 29      F STAT5063      50      11     15     -4     80         N        N
## 30      F STAT5063      50       2      3     -1     10         N        N
##    Introvert d.text.cat U.OSU[index, ]
## 17         3        Neg              Y
## 4          4        Neg              Y
```

```
## 26         5        Neg              N
## 29         3        Neg              N
## 30         4        Neg              Y
```

Question 7.

Take a sample with replacement of size 35 from the data set and compute the 5% trimmed mean for HSClass (trim 5% from each end – see the help file for the mean function). Compare this mean to the 5% trimmed mean of HSClass for the original data. Run the code set.seed(1) before you run the sample function.

```
set.seed(1)
mean(fra$HSClass[sample(1:nrow(fra), 35, replace = T)], trim = 0.05)
```

```
## [1] 309
```

```
mean(Ordered$HSClass[sample(1:nrow(fra), 35, replace = T)], trim = 0.05)
```

```
## [1] 370.9091
```