R Final Exam

Fisher Ankney

December 11th, 2018

Statistics 5193

Question 1.

```
rm(list = ls())
ls()
```

```
## character(0)
```

Question 2.

```
D <- matrix(c(1,0,0,2), nrow = 2)
D[,2]
```

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

Question 3.

```
FRA <- data.frame(read.csv(url('http://jdhabiger.okstate.edu/StudentData.txt'), sep = '\t'))
```

Question 4.

Matrices data storage mode requires the data to be homogeneous, coercing them to lists, while data frame data storage allows for heterogeneous data storage, allowing the variables to remain as factors, integers, numerics, and so on.

Question 5.

```
ordered <- FRA[order(FRA$Gender,FRA$Introvert),]
head(ordered, n = 3)
```

```
##    Gender    Class HSClass TxtSent TxtRec Fbtime Pinterest Snapchat
## 32      F STAT5063     160       5      5     15         N        N
## 9       F STAT2023     200     150    150     20         Y        Y
## 17      F STAT2023     760      20     40     10         Y        Y
##    Introvert
## 32         2
## 9          3
## 17         3
```

Question 6.

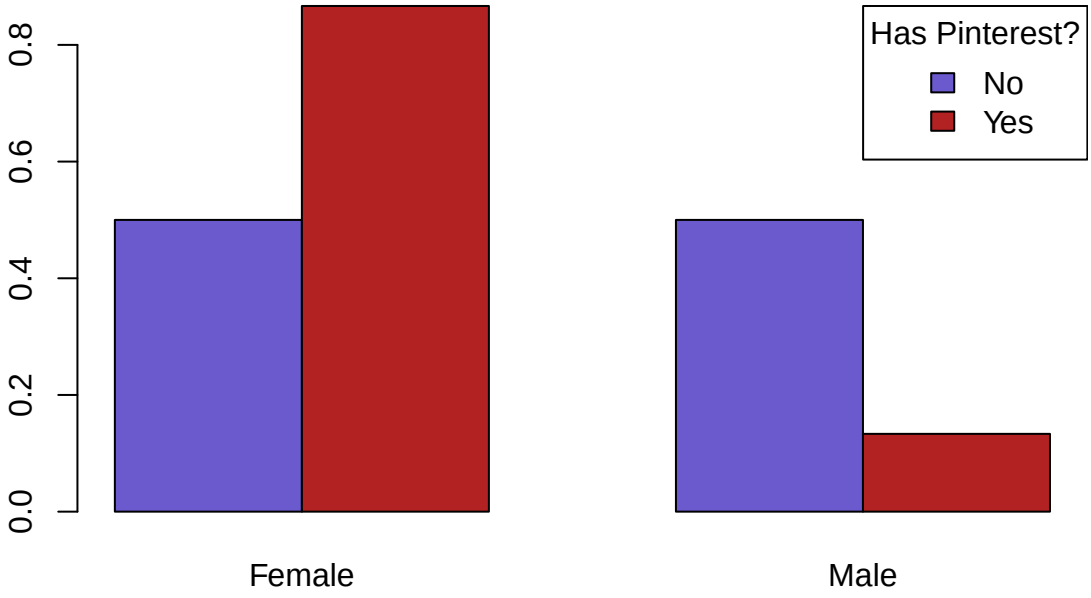
```
pin_gen <- table(FRA$Pinterest, FRA$Gender)
```

Question 7.

```
pin_gen_prop <- prop.table(margin.table(pin_gen, 1:2), 1)
```

Question 8.

```
barplot(pin_gen_prop,
        beside = T,
        names = c('Female', 'Male'),
        col = c('slateblue', 'firebrick', 'slateblue', 'firebrick'))

legend('topright',
        title = 'Has Pinterest?',
        legend = c('No', 'Yes'),
        fill = c('slateblue', 'firebrick'))
```

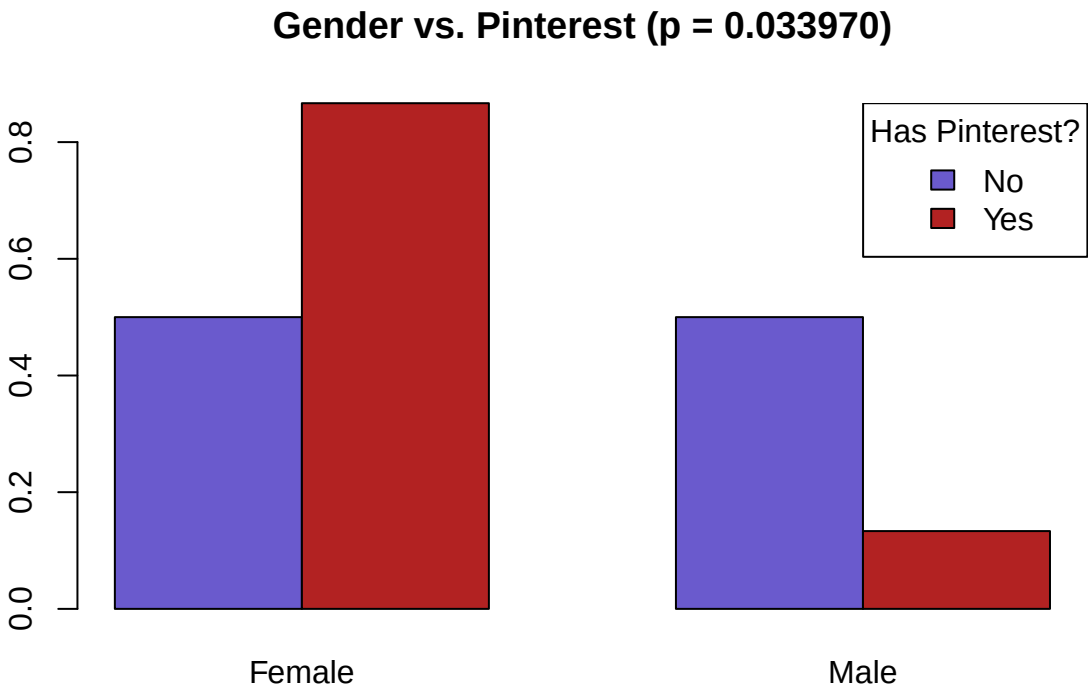


Question 9.

```
fisher.test(pin_gen)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  pin_gen
## p-value = 0.03397
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.01419334 1.01695124
## sample estimates:
## odds ratio
##  0.1623948
```

```
barplot(pin_gen_prop,
        main = 'Gender vs. Pinterest (p = 0.033970)',
        beside = T,
        names = c('Female', 'Male'),
        col = c('slateblue', 'firebrick', 'slateblue', 'firebrick'))

legend('topright',
        title = 'Has Pinterest?',
        legend = c('No', 'Yes'),
        fill = c('slateblue', 'firebrick'))
```



Fisher's exact test yielded that use of Pinterest is not independent of gender, with 0.033970 confidence.

Question 10.

```
resid(chisq.test(pin_gen[,1]))
```

```
##          N          Y
## -0.4423259  0.4423259
```

Slightly more women use pinterest than is expected if gender has no inpact on pinterest use.

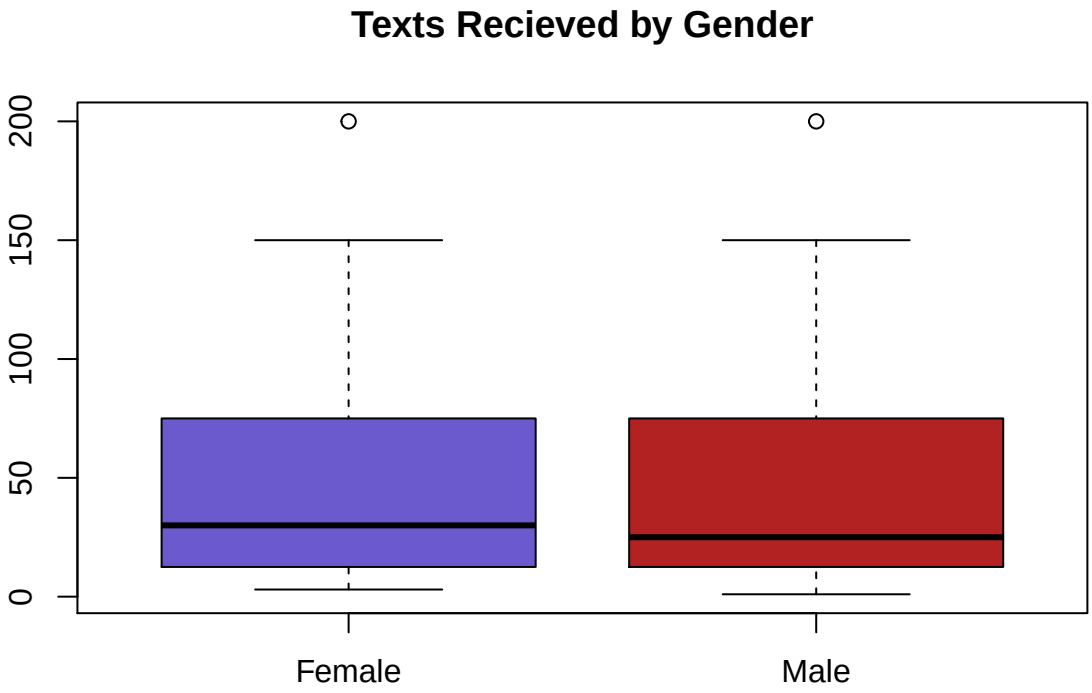
Question 11.

```
ci <- function(mean, std_err) {
  low <- mean - 1.96*std_err
  high <- mean + 1.96*std_err
  cat('The 95% CI is', low, 'and', high)
}

ci(5,2)
```

```
## The 95% CI is 1.08 and 8.92
```

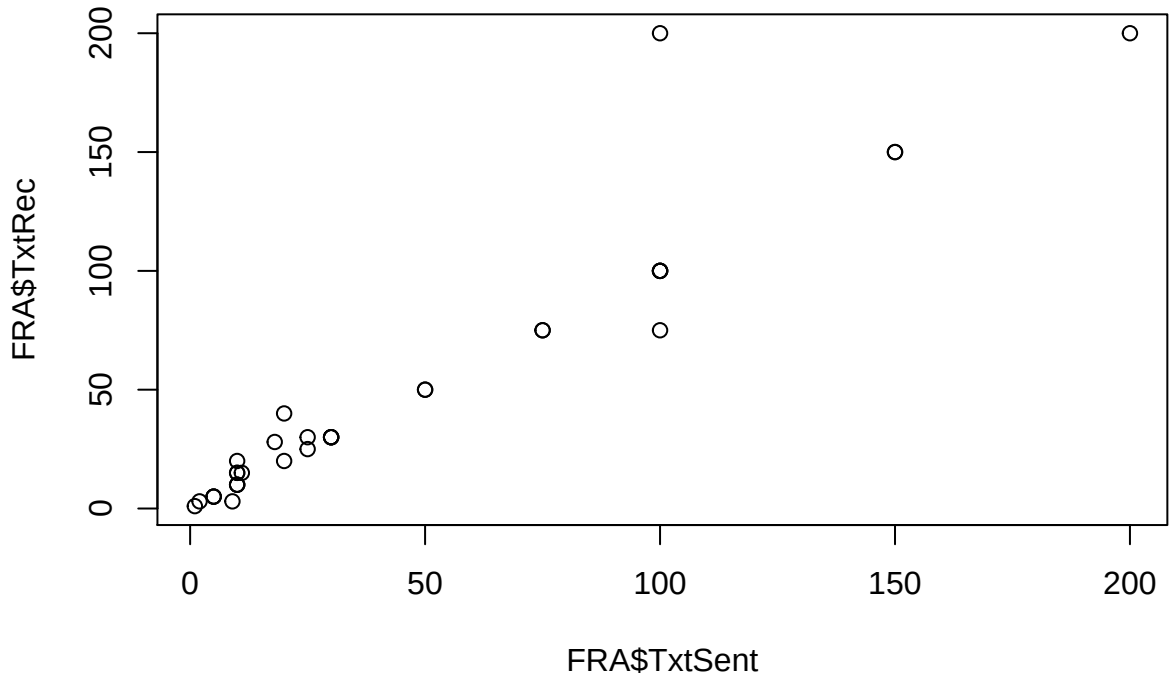
Question 12.

```
boxplot(FRA$TxtRec ~ FRA$Gender,
        col = c('slateblue', 'firebrick'),
        names = c('Female', 'Male'),
        main = 'Texts Recieved by Gender')
```



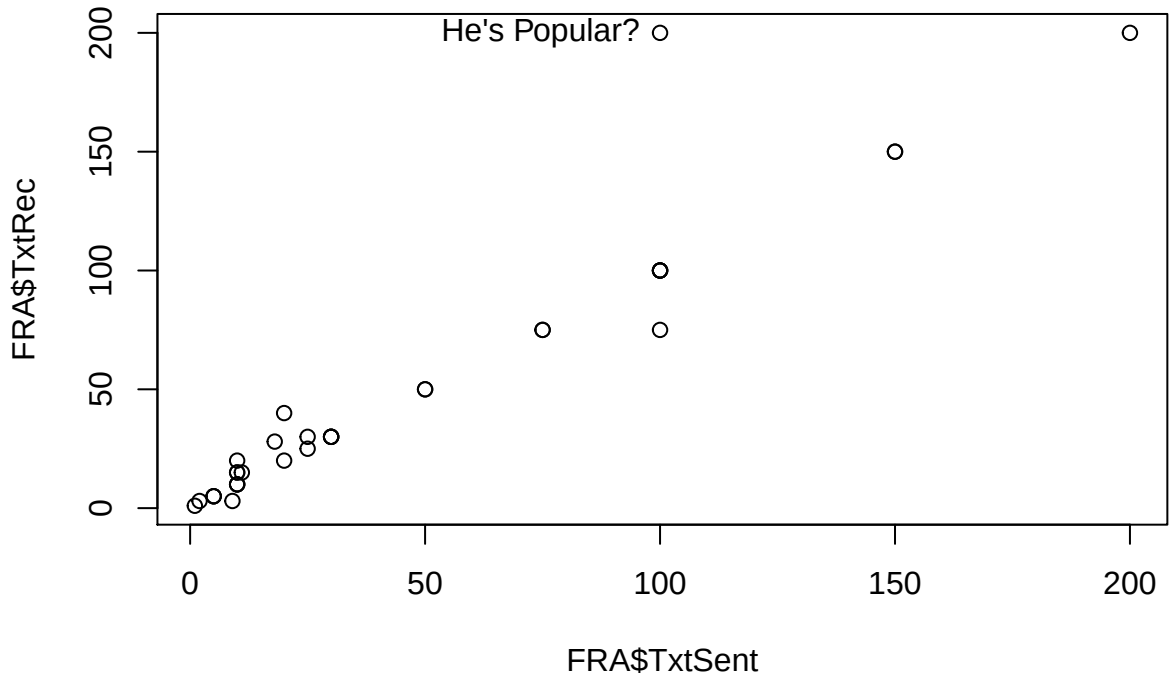
Question 13.

```
plot(FRA$TxtRec ~ FRA$TxtSent)
```



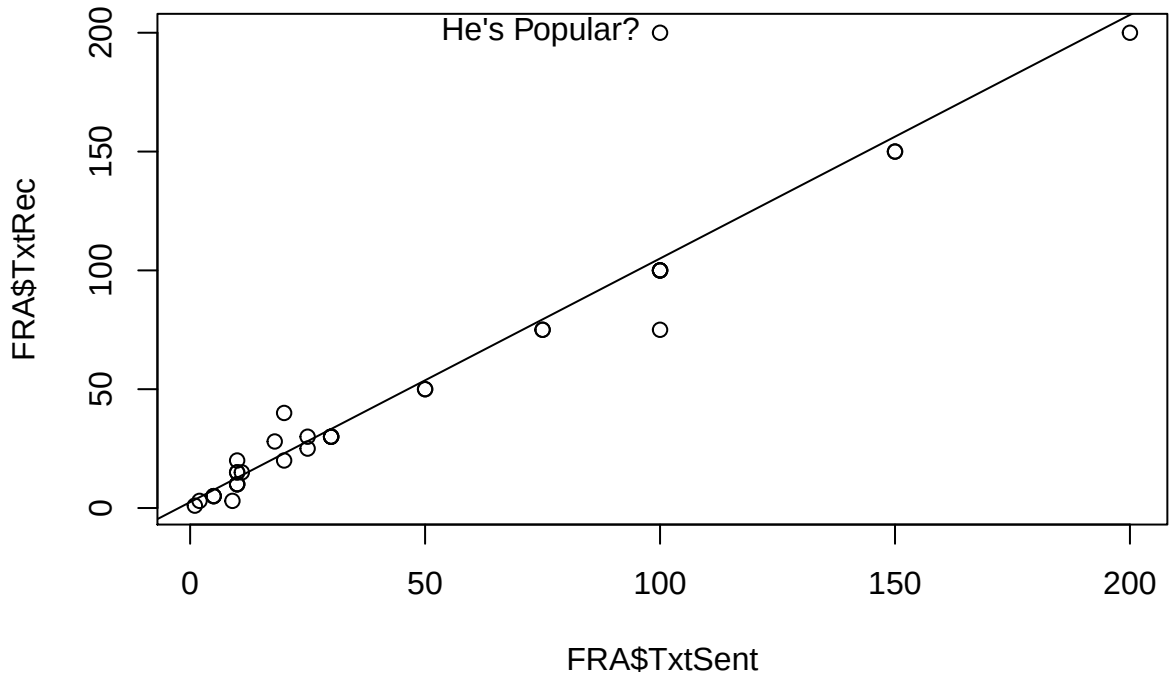
Question 14.

```
plot(FRA$TxtRec ~ FRA$TxtSent)


popular_guy <- FRA[FRA$TxtSent == 100 & FRA$TxtRec == 200,]
text(popular_guy$TxtRec ~ popular_guy$TxtSent,
     label = "He's Popular?",
     pos = 2
)
```



Question 15.

```
plot(FRA$TxtRec ~ FRA$TxtSent)

popular_guy <- FRA[FRA$TxtSent == 100 & FRA$TxtRec == 200,]
text(popular_guy$TxtRec ~ popular_guy$TxtSent,
     label = "He's Popular?",
     pos = 2
)

lines(abline(lm(FRA$TxtRec ~ FRA$TxtSent)))
```



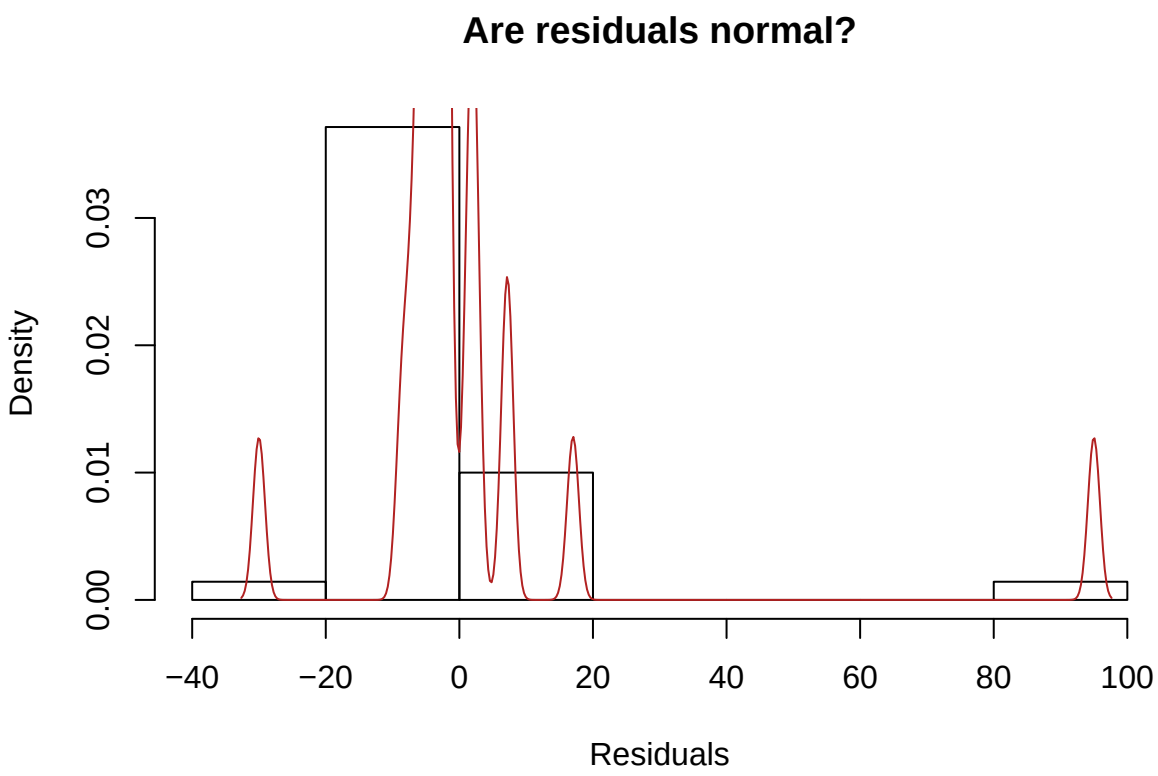
Question 16.

```
text_model <- lm(TxtRec ~ TxtSent, data = FRA)

confint(text_model, 'TxtSent', level = 0.95)
```

```
##              2.5 %   97.5 %
## TxtSent 0.9009659 1.149569
```

Question 17.

```
hist(resid(text_model),
     probability = T,
     main = 'Are residuals normal?',
     xlab = 'Residuals')

lines(density(resid(text_model)), col = 'firebrick')
```



Question 18.

```
t.test(FRA$TxtSent, FRA$TxtRec,
       paired = T)
```

```
##
##  Paired t-test
##
## data:  FRA$TxtSent and FRA$TxtRec
## t = -1.2191, df = 34, p-value = 0.2312
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.829950  2.458522
## sample estimates:
## mean of the differences
##               -3.685714
```

```
sum(FRA$TxtSent)
```

```
## [1] 1676
```

```
sum(FRA$TxtRec)
```

```
## [1] 1805
```

We do not have enough evidence assert that there is a real difference between texts sent and texts received among the population. The p-value is 0.23 and thus, insigificant.