HW 1

Bora Seo

2/26/20

1.

1) 1/3 + 1/4 = 0.5833333

2) 2^{10} + 1 = 1025

3) When f = 440, 1127ln(1+ f/700) = 549.6415

4) When a = -2, b = 4, and c = -4, $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$ = output: 0.7320508

R expressions are as below:

```r
# 1)
first <- (1/3 + 1/4)
print(first)

# 2)
second <- (2^10 + 1)
print(second)

# 3)
f <- 440
third <- 1127*log(1+f/700)
print(third)

# 4)
a <- 2
b <- 4
c <- -4
d <- b^2 - 4*a*c
forth <- (-b+sqrt(d)) / (2*a)
print(forth)
```

```
[1] 0.5833333
[1] 1025
[1] 549.6415
[1] 0.7320508
```

2.

1) *R* in “fourth” was used in 6 times at Klein’s, 13 times in Macy’s, and 16 times in the emphatic condition. The sum of these numbers is 35. I used xtabs function and set up the conditions to earn the required data.

R expressions are as below:

```r
setwd("/Users/velos/Documents/r_project")
nyc_df <- read.csv('NYC.csv')

#Q2_1
result <- ftable(xtabs(~store+word+emphasis, data=nyc_df, subset = r == "1", exclude = c("flooR", "normal")))
result
sum(result)
```

```
                emphasis emphatic
store   word
Klein's fouRth                  6
Macy's  fouRth                 13
Saks    fouRth                 16
[1] 35
```

2) Regardless of use or *r*, “floor” was spoken 104 times by the employees at Klein’s. Among this number, *r* was used 12 times. The percentage is approximately 11.5%.

R expressions are as below:

```r
setwd("/Users/velos/Documents/r_project")
nyc_df <- read.csv('NYC.csv')

#Q2_2

klns_flr <- ftable(xtabs(~word+store, data=nyc_df, exclude = c("Macy's", "Saks", "fouRth")))
klns_flr

r_flr <- ftable(xtabs(~word+store, data=nyc_df, subset = r == "1", exclude = c("Macy's", "Saks", "fouRth")))
r_flr

r_flr[1] / klns_flr[1] * 100
```

```
      store Klein's
word
flooR           104
      store Klein's
word
flooR            12
[1] 11.53846
```

3) Regarding the stretch goal, I looked into the increase or decrease rate in use of *r* between normal and emphatic condition at each department store. I extracted table with xtabs function, and only included where the value or r equals to 1.

1. Macy's:

   "Fourth": Use of *r* decreased by approximately 60.61% in emphatic condition.

   "Floor": Use of *r* decreased by approximately 35.42% in emphatic condition.

2. Klein's:

   "Fourth": Use of *r* doubled in emphatic condition.

   "Floor": Use of *r* increased by 40% in emphatic condition.

3. Sacks:

   "Fourth": There was no change in use or *r*.

   "Floor": Use of *r* increased by approximately 32.26% in emphatic condition.

R expressions are as below:

```{r}
#Q2_stretch
setwd("/Users/velos/Documents/r_project")
nyc_df <- read.csv('NYC.csv')

dep_m_r <- xtabs(~store+word+r+emphasis, data=nyc_df, subset = r == "1", exclude = c("Klein's", "Saks"))
print(m_r <- ftable(dep_m_r))

print(m_r_frth <- ((m_r[4] - m_r[2]) / m_r[4]) * 100)
print(m_r_flr <- ((m_r[3] - m_r[1]) / m_r[3]) * 100 )


dep_k_r <- xtabs(~store+word+r+emphasis, data=nyc_df, subset = r == "1", exclude = c("Macy's", "Saks"))
print(k_r <- ftable(dep_k_r))

print(k_r_frth <- ((k_r[4] - k_r[2]) / k_r[4]) * 100)
print(k_r_flr <- ((k_r[3] - k_r[1]) / k_r[3]) * 100 )

dep_s_r <- xtabs(~store+word+r+emphasis, data=nyc_df, subset = r == "1", exclude = c("Macy's", "Klein's"))
print(s_r <- ftable(dep_s_r))

print(s_r_frth <- ((s_r[4] - s_r[2]) / s_r[4]) * 100)
print(s_r_flr <- ((s_r[3] - s_r[1]) / s_r[3]) * 100 )

```

```
                emphasis emphatic normal
store  word   r
Macy's flooR  1                 31     48
       fouRth 1                 13     33
[1] 60.60606
[1] 35.41667
                  emphasis emphatic normal
store   word   r
Klein's flooR  1                  7      5
        fouRth 1                  6      3
[1] -100
[1] -40
                emphasis emphatic normal
store word   r
Saks  flooR  1                 21     31
      fouRth 1                 16     16
[1] 0
[1] 32.25806
```

The highest *r*-lessness was shown at Macy's when the employees were saying "fourth" between both conditions. In terms of the lowest *r*-lessness, "fourth" at Klein's seems to correspond to the condition since it showed increase of use of *r* unlike other cases. Also, there are certain gaps between the output of each word, however it is difficult to find the coherent tendency.

3.

1) I read VOT.tsv file through read.table function and used quantile function to get the sample quartile of entire VOTs. Q1 = -17,975, Q2 (median) = 13.825, and Q3 = 27.365

2) I assigned a variable to for the VOTs of Spanish speaker and used mean function to get the mean. The output is -24.31306.

3) As in 2, I assigned a variable for the VOTs of English speaker and used sd function to get the standard deviation. The output is 19.86479. For the stretch goal, I used the given formula, and the output here is the same as in previous result.

R expressions are as below:

```
setwd("/Users/velos/Documents/r_project")

#Q3_1
vot_df <- read.table(file='VOT.tsv', sep = '\t', header = TRUE, fill = TRUE)
quantile(vot_df[, "vot"])

#Q3_2
vot_df_sp <- vot_df[vot_df$language == "spanish",]
mean(vot_df_sp[, "vot"])

#Q3_3
vot_df_en <- vot_df[vot_df$language == "english",]
sd(vot_df_en[, "vot"])

#Q3_stretch
val <- vot_df_en[, "vot"]

n <- length(val)
sd_stretch <- sqrt(sum((val - mean(val))^2 / (n - 1)))
sd_stretch
```

```
     0%     25%     50%     75%    100%
-85.290 -17.975  13.825  27.365  82.860
[1] -24.31306
[1] 19.86479
[1] 19.86479
```