Yuying Ren
LING82100: Statistics in Linguistics
Homework 01
Feb/26

1. Arithmetic:

(1)	$\frac{1}{3}+\frac{1}{4}$	> library(MASS) > fractions(1/3 + 1/4) [1] 7/12
(2)	$2^{10}+1$	> 2^10 + 1 [1] 1025
(3)	$1127\mathrm{In}(1+\frac{f}{700})$ $(f = 440)$	> f <- 440 > 1127 * log(1 + f/700) [1] 549.6415
(4)	$\frac{-b\pm\sqrt{b^2-4ac}}{2a}$ (a = 2, b = 4, c = -4)	> f <- function(x, a, b, c) {return (a*x^2+b*x+c)} > re<-uniroot(f, c(0,3), a=2, b=4, c=-4, tol=1e-6) > re$root [1] 0.7320507

2. Categorical data

(1)
```
> setwd("/Users/renyuying/Documents/CUNY Study/Statistics/Stat_HW/stat_hw01")
> df1 = read.csv("rless.csv", header = T)
> dd <- read.csv("rless.csv", header = T)
> summary(dd)
      r          store       emphasis      word
 Min.   :0.0000   Klein's:216   emphatic:271   flooR :347
 1st Qu.:0.0000   Macy's :336   normal  :458   fouRth:382
 Median :0.0000   Saks   :177
 Mean   :0.3155
 3rd Qu.:1.0000
 Max.   :1.0000
> rless <- table(dd$emphasis, dd$word)
> rless

           flooR fouRth
 emphatic   124    147
 normal     223    235
```

The time that employees in the three stores use "r" in "fourth" in emphatic is **147**.

(2)
```
> rless <- table(dd$r, dd$word)
> rless

  flooR fouRth
0   204    295
1   143     87
> rless <- table(dd$r, dd$store, dd$word)
> rless
, ,  = flooR


  Klein's Macy's Saks
0     92    82   30
1     12    79   52

, ,  = fouRth


  Klein's Macy's Saks
0    103   129   63
1      9    46   32

> 12/143
[1] 0.08391608
```

The time that S. Klein's use "r" is **12**, time three stores' eployees use r is **143,** the percentage is **12/143 ≈ 0.084**.

3. Ratio data (I converted the .tsv file to .csv file, but the read.table("VOT.tsv", header = T also worked!)

(1)
```
> dd <- read.csv("VOT.csv", header = T)
> summary(dd)
  participant    language      item      repetition
biling00: 36   english:360   da    : 60   Min.   :1
biling01: 36   spanish:360   de    : 60   1st Qu.:1
biling02: 36                 di    : 60   Median :2
biling03: 36                 dig   : 60   Mean   :2
biling04: 36                 dog   : 60   3rd Qu.:3
biling05: 36                 dug   : 60   Max.   :3
(Other) :504                 (Other):360
      vot
Min.   :-85.29
1st Qu.:-17.98
Median : 13.82
Mean   :  4.06
3rd Qu.: 27.36
Max.   : 82.86
```

The sample quartile of vot is listed above, **25% = -17.98; 50% = 13.82; 75% = 27.36**

(2)
```
> aggregate(dd[, 5], list(dd$language), mean)
  Group.1        x
1 english  32.43242
2 spanish -24.31306
```

The mean of Spanish speakers' VOTs is **-24.31306.**

(3)
```
> aggregate(dd[, 5], list(dd$language), sd)
  Group.1       x
1 english 19.86479
2 spanish 36.41377
```

The standard deviation of English speakers' VOTs is **19.86479.**