Homework 3

Akhil Havaldar (ash2sfp)

Question 1

a)

```
#a)
mean <- 109.8
sd <- 7.4
alpha <- 0.05
print(mean)
```

```
## [1] 109.8
```

```
print(sd)
```

```
## [1] 7.4
```

```
print(alpha)
```

```
## [1] 0.05
```

b)

```
#b)
ztest <- function(size, mean, sd, alpha){
  samp <- rnorm(size, mean = mean, sd = sd)
  mean.samp <- mean(samp)
  zstat <- (mean.samp-mean)/(sd/sqrt(size))
  p <- 2*pnorm(abs(zstat), lower.tail = FALSE)
  ifelse(p<alpha, TRUE, FALSE)
}

print(ztest(23, mean, sd, alpha))
```

```
## [1] FALSE
```

c)

```
#c)
rep <- replicate(10000, ztest(23, mean, sd, alpha))
reject <- sum(rep)/length(rep)
print(reject)
```

```
## [1] 0.0511
```

d

- Theoretically the proportion from part c should be 0.05.

e)

```
#e)
prop <- function(size){
  rep <- replicate(10000, ztest(size, mean, sd, alpha))
  reject <- sum(rep)/length(rep)
}

prop8 <- prop(8)
prop23 <- prop(23)
prop52 <- prop(52)
print(prop8)
```

```
## [1] 0.0479
```

```
print(prop23)
```

```
## [1] 0.0548
```

```
print(prop52)
```

```
## [1] 0.0473
```

f)

```r
#f)
lapply(3:52, prop)
```

```
## [[1]]
## [1] 0.0497
##
## [[2]]
## [1] 0.0494
##
## [[3]]
## [1] 0.0507
##
## [[4]]
## [1] 0.0485
##
## [[5]]
## [1] 0.0495
##
## [[6]]
## [1] 0.0522
##
## [[7]]
## [1] 0.0473
##
## [[8]]
## [1] 0.05
##
## [[9]]
## [1] 0.0486
##
## [[10]]
## [1] 0.0495
##
## [[11]]
## [1] 0.0515
##
## [[12]]
## [1] 0.0504
##
## [[13]]
## [1] 0.0508
##
## [[14]]
## [1] 0.0505
```

```
##
## [[15]]
## [1] 0.0504
##
## [[16]]
## [1] 0.0489
##
## [[17]]
## [1] 0.0523
##
## [[18]]
## [1] 0.0474
##
## [[19]]
## [1] 0.0453
##
## [[20]]
## [1] 0.05
##
## [[21]]
## [1] 0.0473
##
## [[22]]
## [1] 0.0504
##
## [[23]]
## [1] 0.0504
##
## [[24]]
## [1] 0.0522
##
## [[25]]
## [1] 0.0525
##
## [[26]]
## [1] 0.0476
##
## [[27]]
## [1] 0.0492
##
## [[28]]
## [1] 0.0477
##
## [[29]]
## [1] 0.0485
```

```
##
## [[30]]
## [1] 0.0519
##
## [[31]]
## [1] 0.0513
##
## [[32]]
## [1] 0.0492
##
## [[33]]
## [1] 0.0491
##
## [[34]]
## [1] 0.0506
##
## [[35]]
## [1] 0.0515
##
## [[36]]
## [1] 0.0525
##
## [[37]]
## [1] 0.0489
##
## [[38]]
## [1] 0.0468
##
## [[39]]
## [1] 0.0522
##
## [[40]]
## [1] 0.0521
##
## [[41]]
## [1] 0.0479
##
## [[42]]
## [1] 0.0497
##
## [[43]]
## [1] 0.0442
##
## [[44]]
## [1] 0.0548
```

```
##
## [[45]]
## [1] 0.0495
##
## [[46]]
## [1] 0.0529
##
## [[47]]
## [1] 0.0497
##
## [[48]]
## [1] 0.0513
##
## [[49]]
## [1] 0.047
##
## [[50]]
## [1] 0.051
```

g

- The sample size does not seem to have an impact on the results. With every sample size, the proportion seems to be between [0.04, 0.06].

Question 2

a)

```
#a)

nym2021 <- read.table('nym2021.txt', header=TRUE)
head(nym2021)
```

```
##   Sex Age Place DivPlace    DIV DivAge   Time BostonQualifier
## 1   M  35  1593      269 M35-39  35-39 198.90               N
## 2   M  28   544       96 M25-29  25-29 178.70               Y
## 3   M  32  2296      399 M30-34  30-34 206.63               N
## 4   M  34  1192      239 M30-34  30-34 191.90               N
## 5   F  26    64        4 F25-29  25-29 154.85               Y
## 6   M  28   536       94 M25-29  25-29 178.52               Y
##   HomeStateOrCountry
## 1                 NY
```

```
## 2                  NY
## 3                  NY
## 4                  NY
## 5                 MEX
## 6                 POL
```

b)

```
#b)
print(length(nym2021$Time))
```

```
## [1] 275
```

c)

```
#c)
temp <- nym2021[nchar(as.character(nym2021$HomeStateOrCountry)) == 2, ]
print(length(temp$HomeStateOrCountry))
```

```
## [1] 210
```

d)

```
#d)
nym2021$country <- ifelse(nchar(as.character(nym2021$HomeStateOrCountry)) == 2,
                          "USA", nym2021$HomeStateOrCountry)
table(nym2021$country)
```

```
##
## AUS BRA BTN CAN COL CRI DEU DNK DOM EGY ESP ETH FRA GBR HUN IND IRL ISR ITA JPN
##   1   1   1   8   1   1   3   1   2   1   2   2   7   7   1   2   2   1   1   4
## MEX NLD NOR PER POL SWE UKR USA ZAF
##   8   1   1   1   1   1   2 210   1
```

e)

```r
#e)
print(length(unique(nym2021$country)))
```

```
## [1] 29
```

f)

```r
#f)
print(c("Youngest:", min(nym2021$Age)))
```

```
## [1] "Youngest:" "20"
```

```r
print(c("Oldest:", max(nym2021$Age)))
```

```
## [1] "Oldest:" "62"
```

g)

```r
#g)
temp2 <- nym2021[nym2021$Time == max(nym2021$Time),]
print(c("Age Slowest:", temp2$Age))
```

```
## [1] "Age Slowest:" "27"
```

```r
temp3 <- nym2021[nym2021$Time == min(nym2021$Time),]
print(c("Age Fastest:", temp3$Age))
```

```
## [1] "Age Fastest:" "38"
```

h)

```r
#h)
temp4 <- nym2021[which(nym2021$DivPlace<=25),]
print(length(temp4$DivPlace))
```

```
## [1] 56
```

i)

```
#i)
print(sort(unique(temp4$DIV)))
```

```
##  [1] "F20-24" "F25-29" "F30-34" "F35-39" "F40-44" "F45-49" "F50-54" "F60-64"
##  [9] "M20-24" "M25-29" "M30-34" "M35-39" "M40-44" "M45-49" "M50-54" "M55-59"
## [17] "M60-64"
```

j)

```
#j)
temp5 <- nym2021[which(nym2021$DivPlace<=10),]
print(temp5)
```

```
##     Sex Age Place DivPlace    DIV DivAge   Time BostonQualifier
## 5     F  26    64        4 F25-29  25-29 154.85               Y
## 10    F  50  2080       10 F50-54  50-54 204.57               Y
## 27    M  50   243        2 M50-54  50-54 169.42               Y
## 41    M  52   517        9 M50-54  50-54 178.07               Y
## 44    M  36    58        8 M35-39  35-39 153.55               Y
## 58    F  27    26        2 F25-29  25-29 144.70               Y
## 106   M  57   508        5 M55-59  55-59 177.90               Y
## 110   F  36   242        8 F35-39  35-39 169.42               Y
## 111   M  60  1519        4 M60-64  60-64 197.77               Y
## 126   F  28   224        9 F25-29  25-29 168.58               Y
## 135   M  29    15        6 M25-29  25-29 137.42               Y
## 143   M  25    13        5 M25-29  25-29 136.65               Y
## 147   F  41    70        3 F40-44  40-44 155.90               Y
## 161   M  24    75        3 M20-24  20-24 156.65               Y
## 184   F  62  2076        1 F60-64  60-64 204.53               Y
## 196   F  22  1445       10 F20-24  20-24 196.63               Y
## 203   M  50   527       10 M50-54  50-54 178.40               Y
## 205   M  33    10        4 M30-34  30-34 134.10               Y
## 216   M  57   578        7 M55-59  55-59 179.20               Y
## 224   F  30    94        5 F30-34  30-34 158.90               Y
## 226   M  38     4        1 M35-39  35-39 131.25               Y
## 228   F  30    24        2 F30-34  30-34 142.87               Y
## 234   F  33   250       10 F30-34  30-34 169.57               Y
## 249   M  45    81        1 M45-49  45-49 157.45               Y
## 274   M  57   513        6 M55-59  55-59 177.97               Y
```

```
##     HomeStateOrCountry country
## 5                  MEX     MEX
## 10                  NY     USA
## 27                  CO     USA
## 41                  CT     USA
## 44                  CA     USA
## 58                  AZ     USA
## 106                 NY     USA
## 110                CAN     CAN
## 111                CAN     CAN
## 126                 NJ     USA
## 135                DNK     DNK
## 143                JPN     JPN
## 147                SWE     SWE
## 161                 NJ     USA
## 184                 FL     USA
## 196                 NY     USA
## 203                JPN     JPN
## 205                 UT     USA
## 216                MEX     MEX
## 224                 AZ     USA
## 226                 CO     USA
## 228                ETH     ETH
## 234                ITA     ITA
## 249                ZAF     ZAF
## 274                 GA     USA
```

k)

```
#k)
print(tapply(nym2021$Age, nym2021$BostonQualifier, mean))
```

```
##        N        Y
## 34.20769 37.49655
```