Homework 3

Will Scheib

Problem 1

Part a

```
mean_norm <- 99.5
sd_norm <- 4.8
alpha_norm <- 0.05
```

Part b

```
sample_norm <- function(n) {
  sub <- rnorm(n, mean_norm, sd_norm)
  samp_mean <- mean(sub)
  z_stat <- (mean_norm-samp_mean)/(sd_norm/sqrt(n))
  2*pnorm(-abs(z_stat)) < 0.05
}

sample_norm(27)
```

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

Part c

```
mean(replicate(10000, sample_norm(27)))
```

```
## [1] 0.0488
```

Part d

Theoretically, the proportion should be 0.05.

Part e

```
rep_norm <- function(n) {
  mean(replicate(10000, sample_norm(n)))
}

sapply(c(9, 27, 51), rep_norm)
```

```
## [1] 0.0501 0.0518 0.0501
```

Part f

```
sapply(3:51, rep_norm)
```

```
##  [1] 0.0480 0.0505 0.0506 0.0452 0.0466 0.0485 0.0543 0.0471 0.0483 0.0471
## [11] 0.0480 0.0496 0.0476 0.0488 0.0478 0.0517 0.0502 0.0543 0.0524 0.0483
## [21] 0.0475 0.0556 0.0532 0.0477 0.0478 0.0502 0.0487 0.0498 0.0474 0.0475
## [31] 0.0512 0.0517 0.0479 0.0475 0.0519 0.0498 0.0502 0.0546 0.0510 0.0495
## [41] 0.0495 0.0500 0.0483 0.0465 0.0455 0.0490 0.0523 0.0545 0.0510
```

Part g

They don't change in any consistent way. Sample size does not seem to affect the results.

Problem 2

Part a

```
nym2019 <- read.table("data/nym2019.txt", header=TRUE)
head(nym2019)
```

```
##   Sex Age Place DivPlace    DIV DivAge   Time BostonQualifier
## 1   M  38  5824      947 M35-39  35-39 208.80               N
## 2   M  44 18719     2314 M40-44  40-44 248.27               N
## 3   M  56 14716      609 M55-59  55-59 237.72               Y
## 4   M  48 11240     1327 M45-49  45-49 228.72               N
## 5   M  44  1572      248 M40-44  40-44 180.15               N
## 6   M  28   245       64 M25-29  25-29 161.42               Y
##   HomeStateOrCountry
## 1                 NY
## 2                 NC
## 3                 NY
## 4                 NJ
## 5                ESP
## 6                 NY
```

Part b

```
nrow(nym2019)
```

```
## [1] 400
```

Part c

```
nym2019[nchar(nym2019$HomeStateOrCountry) == 2,"HomeStateOrCountry"] <- "USA"
nrow(nym2019[nym2019$HomeStateOrCountry == "USA",])
```

```
## [1] 191
```

Part d

```
table(nym2019$HomeStateOrCountry)
```

```
##
## AND ARG AUS AUT BEL BRA CAN CHN COL CZE DEN ECU ESA ESP ETH FRA GBR GER GUA HKG
##   1   1  10   2   2   4  15   6   3   1   4   2   1  13   6  25  20  10   1   2
## HUN INA IRL ITA JPN KEN MEX NCA NED NOR NZL PER PHI POL POR RSA RUS SIN SRI SUI
##   1   1   5  17   4   2   6   1   9   3   1   2   1   4   2   1   1   1   1   5
## SWE THA TPE UGA UKR USA VEN
##   6   1   1   1   1 191   2
```

Part e

```
length(unique(nym2019$HomeStateOrCountry))
```

```
## [1] 47
```

Part f

```
young_old <- c(youngest=min(nym2019$Age), oldest=max(nym2019$Age))
young_old
```

```
## youngest   oldest
##       21       71
```

Part g

```
fast_slow <- c(
  fastest=nym2019[nym2019$Time==min(nym2019$Time),"Age"],
  slowest=nym2019[nym2019$Time==max(nym2019$Time),"Age"]
)
fast_slow
```

```
## fastest slowest
##      23      41
```

Part h

```
nrow(nym2019[nym2019$DivPlace<=20,])
```

```
## [1] 31
```

Part i

```
sort(unique(nym2019[nym2019$DivPlace <= 20,"DIV"]))
```

```
##  [1] "F20-24" "F25-29" "F30-34" "F35-39" "F40-44" "M20-24" "M25-29" "M30-34"
##  [9] "M35-39" "M40-44" "M45-49" "M50-54" "M70-74"
```

Part j

```
nym2019[nym2019$DivPlace <= 5,]
```

```
##     Sex Age Place DivPlace    DIV DivAge   Time BostonQualifier
## 13    M  70  6929        4 M70-74  70-74 213.37               Y
## 56    M  71  9278        5 M70-74  70-74 222.43               N
## 63    M  40    25        2 M40-44  40-44 139.68               N
## 126   M  38    11        1 M35-39  35-39 132.95               Y
## 137   F  41    74        3 F40-44  40-44 150.20               N
## 159   M  23     5        1 M20-24  20-24 130.65               Y
## 172   M  46    91        3 M45-49  45-49 153.05               N
## 281   F  24   265        1 F20-24  20-24 162.35               Y
## 389   F  25    39        2 F25-29  25-29 145.85               Y
##     HomeStateOrCountry
## 13                 CHN
## 56                 USA
## 63                 SWE
## 126                GER
## 137                USA
## 159                ETH
## 172                USA
## 281                ETH
## 389                ETH
```

Part k

```
tapply(nym2019$Age, nym2019$BostonQualifier, mean)
```

```
##        N        Y
## 39.25234 38.95699
```