Homework 3

Carlos F Revilla

Problem 1

a)

```
mean<-109.8
sd<-7.4
print(mean)
```

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

```
print(sd)
```

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

b)

```
ztest<-function(sample_size){
sample<-rnorm(sample_size,mean,sd)
z<-(mean(sample)-mean)/(sd/sqrt(sample_size))
p<-pnorm(abs(z),lower.tail=FALSE)
p<0.05
}
ztest(23)
```

```
## [1] TRUE
```

```
print(ztest(23))
```

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

c)

```
# Execute function 10,000 times using a sample size of 23
results <- replicate(10000, ztest(23))

# Calculate proportion of tests that reject the null hypothesis and print
print(mean(results))
```

```
## [1] 0.1048
```

d)

```
#Theoretically,the proportion should be around 5% as that is the significance level
```

e)

```
proportion_rejecting_null <- function(mean, sd, sample_size, num_tests) {
  # Execute z-test function num_tests times using a sample size of sample_size
  results <- replicate(num_tests, ztest(sample_size))
  return(mean(results))
}

proportion_rejecting_null(mean, sd, 8, 10000)
```

```
## [1] 0.1007
```

```
proportion_rejecting_null(mean, sd, 23, 10000)
```

```
## [1] 0.104
```

```
proportion_rejecting_null(mean, sd, 52, 10000)
```

```
## [1] 0.1035
```

f)

```
# Generate sequence of sample sizes from 3 to 52
sample_sizes <- 3:52

# Execute function for each sample size in sequence
proportions <- lapply(sample_sizes, function(x) proportion_rejecting_null(mean, sd, x, 1

# View results
proportions
```

```
## [[1]]
## [1] 0.0963
##
## [[2]]
## [1] 0.0993
##
## [[3]]
## [1] 0.1041
##
```

```
## [[4]]
## [1] 0.1026
##
## [[5]]
## [1] 0.102
##
## [[6]]
## [1] 0.0979
##
## [[7]]
## [1] 0.1086
##
## [[8]]
## [1] 0.0997
##
## [[9]]
## [1] 0.0961
##
## [[10]]
## [1] 0.102
##
## [[11]]
## [1] 0.0984
##
## [[12]]
## [1] 0.1007
##
## [[13]]
## [1] 0.1011
##
## [[14]]
## [1] 0.1032
##
## [[15]]
## [1] 0.0944
##
## [[16]]
## [1] 0.0992
##
## [[17]]
## [1] 0.1028
##
## [[18]]
## [1] 0.0981
##
```

```
## [[19]]
## [1] 0.1009
##
## [[20]]
## [1] 0.1007
##
## [[21]]
## [1] 0.0982
##
## [[22]]
## [1] 0.0984
##
## [[23]]
## [1] 0.0984
##
## [[24]]
## [1] 0.1006
##
## [[25]]
## [1] 0.1013
##
## [[26]]
## [1] 0.1022
##
## [[27]]
## [1] 0.0985
##
## [[28]]
## [1] 0.1004
##
## [[29]]
## [1] 0.1006
##
## [[30]]
## [1] 0.096
##
## [[31]]
## [1] 0.0982
##
## [[32]]
## [1] 0.0999
##
## [[33]]
## [1] 0.0976
##
```

```
## [[34]]
## [1] 0.1029
##
## [[35]]
## [1] 0.0998
##
## [[36]]
## [1] 0.098
##
## [[37]]
## [1] 0.0994
##
## [[38]]
## [1] 0.1004
##
## [[39]]
## [1] 0.0972
##
## [[40]]
## [1] 0.0978
##
## [[41]]
## [1] 0.1015
##
## [[42]]
## [1] 0.1003
##
## [[43]]
## [1] 0.1072
##
## [[44]]
## [1] 0.0993
##
## [[45]]
## [1] 0.0932
##
## [[46]]
## [1] 0.1073
##
## [[47]]
## [1] 0.1063
##
## [[48]]
## [1] 0.0966
##
```

```
## [[49]]
## [1] 0.0977
##
## [[50]]
## [1] 0.0991
```

g)

```
#proportion of tests that reject the null hypothesis generally decreases as the sample
```

Problem 2

a)

```
nym2021<-read.csv("nym2021.txt")
head(nym2021)
```

```
##   Sex.Age.Place.DivPlace.DIV.DivAge.Time.BostonQualifier.HomeStateOrCountry
## 1                                M\t35\t1593\t269\tM35-39\t35-39\t198.9\tN\tNY
## 2                                  M\t28\t544\t96\tM25-29\t25-29\t178.7\tY\tNY
## 3                               M\t32\t2296\t399\tM30-34\t30-34\t206.63\tN\tNY
## 4                                M\t34\t1192\t239\tM30-34\t30-34\t191.9\tN\tNY
## 5                                  F\t26\t64\t4\tF25-29\t25-29\t154.85\tY\tMEX
## 6                                M\t28\t536\t94\tM25-29\t25-29\t178.52\tY\tPOL
```

b)

```
print(nrow(nym2021))
```

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

c)

```
print(sum(nym2021$Home.Country=="UnitedStates"))
```

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

d)

```
print(table(nym2021$Home.Country))
```

```
## < table of extent 0 >
```

e)

```
print(length(unique(nym2021$Home.Country)))
```

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

f)

```
print(min(nym2021$Age))
```

```
## Warning in min(nym2021$Age): no non-missing arguments to min; returning Inf
## [1] Inf
```

```
print(max(nym2021$Age))
```

```
## Warning in max(nym2021$Age): no non-missing arguments to max; returning -Inf
## [1] -Inf
```

g)

```
print(nym2021[which.min(nym2021$Finish.Time),"Age"])
```

```
## NULL
```

```
print(nym2021[which.max(nym2021$Finish.Time),"Age"])
```

```
## NULL
```

h)

```
print(sum(nym2021$Overall.Rank<=25))
```

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

i)

```
print(unique(nym2021[nym2021$Overall.Rank<=25,"Division"]))
```

```
## NULL
```

j)

```
print(nym2021[nym2021$Overall.Rank<=10,])
```

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

k)

```
print(mean(nym2021[nym2021$Qualified.for.Boston.Marathon=="Yes","Age"]))
```

```
## Warning in mean.default(nym2021[nym2021$Qualified.for.Boston.Marathon == :
## argument is not numeric or logical: returning NA
```

```
## [1] NA
```

```
print(mean(nym2021[nym2021$Qualified.for.Boston.Marathon=="No","Age"]))
```

```
## Warning in mean.default(nym2021[nym2021$Qualified.for.Boston.Marathon == :
## argument is not numeric or logical: returning NA
```

```
## [1] NA
```