> ### Part 1) Central Limit Theorem

> boston <- read.csv(

+ "https://people.bu.edu/kalathur/datasets/bostonCityEarnings.csv",

+ colClasses = c("character", "character", "character", "integer", "character"))

> head(boston)

NAME Department Title Earnings

1 Abadi,Kidani A Assessing Department Property Utilization Officer 71014

2 Abajue Umeh,Kenechukwu Stephanie BPS Boston Arts Academy Teacher 69087

3 Abasciano,Joseph Boston Police Department Police Officer 151022

4 Abban,Christopher John Boston Fire Department Fire Fighter 153424

5 Abberton,James P Public Works Department Maint Mech (Carpenter)## 58488

6 Abbott,John R. BPS Snowden International Hi Teacher 107720

ZipCode

1 02118

2 02150

3 02132

4 02132

5 02127

6 02445

> ##a)

> hist(boston$Earnings, xlim = c(40000,400000),breaks= 18)

> mean(boston$Earnings)

[1] 108680.9

> sd(boston$Earnings)

[1] 50474.7

> ## Question b)

> set.seed(948)

> samples <- 5000

> sample.size <- 10

> sampleb <- numeric(samples)

> for (i in 1:samples){

+ sampleb[i] <- mean(sample(boston$Earnings,sample.size, replace = FALSE))

+ }

> hist(sampleb)

> mean(sampleb)

[1] 108794.9

> sd(sampleb)

[1] 15934.45

> ## Question c)

> set.seed(948)

> samples <- 5000

> sample.size <- 40

> samplec <- numeric(samples)

> for (i in 1:samples){

+ samplec[i] <- mean(sample(boston$Earnings,sample.size, replace = FALSE))

+ }

> hist(samplec)

> mean(samplec)

[1] 108599.6

> sd(samplec)

[1] 7972.063

> set.seed(948)

> r <- 3

> p <- 0.5

> samples <- 5000

> xbar <- numeric(samples)

> for( i in 1:samples){

+ xbar[i] <- rnbinom(samples,size = r, prob = p)

+ }

There were 50 or more warnings (use warnings() to see the first 50)

> par(mfrow = c(1,1))

> barplot(prop.table(table(xbar)))

> mean(xbar)

[1] 2.9758

> sd(xbar)

[1] 2.440247

> ## Question b)

> samples <- 1000

> sampled <- numeric(samples)

> par(mfrow = c(2,2))

> for (size in c(10, 20, 30, 40)){

+ for ( i in 1:samples){

+ sampled[i] <- mean(sample(xbar,size, replace = FALSE))

+ }

+ hist(sampled,

+ main = paste("Sample Size =", size),prob = TRUE)

+ cat("Sample Size = ", size, " Mean = ", mean(sampled),

+ " SD = ", sd(sampled), "\n")

+ }

Sample Size = 10 Mean = 2.9806 SD = 0.7954095

Sample Size = 20 Mean = 2.96985 SD = 0.5497187

Sample Size = 30 Mean = 2.9754 SD = 0.4604782

Sample Size = 40 Mean = 2.9838 SD = 0.3765991

> par(mfrow = c(1,1))

> mean(sampled)

[1] 2.9838

> sd(sampled)

[1] 0.3765991

> library(prob)

Loading required package: combinat

Attaching package: ‘combinat’

The following object is masked from ‘package:utils’:

combn

Loading required package: fAsianOptions

Loading required package: timeDate

Loading required package: timeSeries

Loading required package: fBasics

Loading required package: fOptions

Attaching package: ‘prob’

The following objects are masked from ‘package:base’:

intersect, setdiff, union

> ### Part 3) Sampling

> a <- as.data.frame(sort(table(boston$Department),decreasing = TRUE))

> b <- a[1:5,1]

> c <- subset(boston, boston$Department %in% b )

> rownames(c) <- 1:nrow(c)

> head(c)

NAME Department Title Earnings ZipCode

1 Abasciano,Joseph Boston Police Department Police Officer 151022 02132

2 Abban,Christopher John Boston Fire Department Fire Fighter 153424 02132

3 Abdul-Aziz,Ramadani Boston Police Department Police Officer 176117 02119

4 Abrahamson,Patrick Olaf Boston Police Department Police Officer 106423 02135

5 Abreu,Cesar Boston Police Department Police Officer 125081 02136

6 Abreu,Gabriel Boston Police Department Police Officer 110116 02136

> sample.size <- 50

> library(sampling)

> ## Question a)

> set.seed(948)

> s <- srswor(70, nrow(c))

> sample.1 <- c[s != 0, ]

> table(sample.1$Department)

Boston Fire Department Boston Police Department Boston Public Library BPS Facility Management

18 38 6 4

BPS Special Education

4

> ## Question b)

> set.seed(948)

> N <- nrow(c)

> k <- ceiling(N / sample.size)

> k

[1] 117

> r <- sample(k, 1)

> r

[1] 41

> s <- seq(r, by = k, length = sample.size)

> sample.2 <- c[s, ]

> table(sample.2$Department)

Boston Fire Department Boston Police Department Boston Public Library BPS Facility Management

11 25 8 4

BPS Special Education

2

> ## Question c)

> set.seed(948)

> pik <- inclusionprobabilities(

+ c$Earnings, sample.size)

> length(pik)

[1] 5814

> sum(pik)

[1] 50

> s <- UPsystematic(pik)

> sample.3 <- c[s != 0, ]

> table(sample.3$Department)

Boston Fire Department Boston Police Department Boston Public Library BPS Facility Management

18 28 1 2

BPS Special Education

1

> ## Question d)

> set.seed(948)

> order.index <- order(c$Department)

> data <- c[order.index, ]

> st <- sampling::strata(data, stratanames = c("Department"),

+ size = c(14,22,8,4,11,5,6) ,

+ method = "srswor")

> sample.5 <- getdata(data, st)

> table(sample.5$Department)

Boston Fire Department Boston Police Department Boston Public Library BPS Facility Management

14 22 8 4

BPS Special Education

11

> ## Question e)

> mean(sample.1$Earnings)

[1] 140085.7

> mean(sample.2$Earnings)

[1] 133591.4

> mean(sample.3$Earnings)

[1] 159511.7

> mean(sample.5$Earnings)

[1] 125765.9

>