> library( readr )

> population\_data <- read.csv("./iris.csv", header = TRUE )

> #View(population\_data)

>

> #Q1(a)

> #SRSWOR 30

> n = 30 # sample size

> data\_i <- population\_data$X

> N = length(data\_i) # population size

> sample\_i <- sample(data\_i,n,replace=FALSE)

> print(sample\_i)

[1] 63 37 19 32 3 145 33 136 30 142 130 150 56 90 77 114 39 133 8 88 14 41 101 120 13 103

[27] 91 115 79 35

> sample\_sepal\_length <- c(-1,n)

> sample\_sepal\_width <- c(-1,n)

> sample\_petal\_length <- c(-1,n)

> sample\_petal\_width <- c(-1,n)

> for (i in 1:n){

+ sample\_sepal\_length[i] <- population\_data$Sepal.Length[sample\_i[i]]

+ sample\_sepal\_width[i] <- population\_data$Sepal.Width[sam .... [TRUNCATED]

> sample\_sepal\_length\_mean = mean(sample\_sepal\_length)

> std\_err\_sepal\_length\_mean = sqrt(1-n/N)\*sd(sample\_sepal\_length)/sqrt(n)

> sprintf("Estimate for population sepal length mean: %f",sample\_sepal\_length\_mean)

[1] "Estimate for population sepal length mean: 5.770000"

> sprintf("Standard Error for estimate of population sepal length mean: %f",std\_err\_sepal\_length\_mean)

[1] "Standard Error for estimate of population sepal length mean: 0.141529"

> sample\_sepal\_width\_mean = mean(sample\_sepal\_width)

> std\_err\_sepal\_width\_mean = sqrt(1-n/N)\*sd(sample\_sepal\_width)/sqrt(n)

> sprintf("Estimate for population sepal width mean: %f",sample\_sepal\_width\_mean)

[1] "Estimate for population sepal width mean: 3.010000"

> sprintf("Standard Error for estimate of population sepal width mean: %f",std\_err\_sepal\_width\_mean)

[1] "Standard Error for estimate of population sepal width mean: 0.071419"

> sample\_petal\_length\_mean = mean(sample\_petal\_length)

> std\_err\_petal\_length\_mean = sqrt(1-n/N)\*sd(sample\_petal\_length)/sqrt(n)

> sprintf("Estimate for population petal length mean: %f",sample\_petal\_length\_mean)

[1] "Estimate for population petal length mean: 3.600000"

> sprintf("Standard Error for estimate of population petal length mean: %f",std\_err\_petal\_length\_mean)

[1] "Standard Error for estimate of population petal length mean: 0.308321"

> sample\_petal\_width\_mean = mean(sample\_petal\_width)

> std\_err\_petal\_width\_mean = sqrt(1-n/N)\*sd(sample\_petal\_width)/sqrt(n)

> sprintf("Estimate for population petal width mean: %f",sample\_petal\_width\_mean)

[1] "Estimate for population petal width mean: 1.156667"

> sprintf("Standard Error for estimate of population petal width mean: %f",std\_err\_petal\_width\_mean)

[1] "Standard Error for estimate of population petal width mean: 0.143306"

> #Q1(b) stratified sampling

> n = 30

> data\_i <- population\_data$X

> N = length(data\_i) # population size

> setosa\_N = 0

> versicolor\_N = 0

> virginica\_N = 0

> for (i in 1:N){

+ type <- population\_data$Species[i]

+ if(type == 'setosa'){

+ setosa\_N = setosa\_N+1

+ }else if(type == 'versicolor'){

+ .... [TRUNCATED]

> stopifnot(setosa\_N+virginica\_N+versicolor\_N == N)

> setosa\_population\_i <- c(-1,setosa\_N)

> x <- 1

> versicolor\_population\_i <- c(-1,versicolor\_N)

> y <- 1

> virginica\_population\_i <- c(-1,virginica\_N)

> z <- 1

> for (i in 1:N){

+ type <- population\_data$Species[i]

+ if(type == 'setosa'){

+ setosa\_population\_i[x] = population\_data$X[i]

+ x = x+1

+ .... [TRUNCATED]

> setosa\_n = setosa\_N/N\*n

> virginica\_n = virginica\_N/N\*n

> versicolor\_n = versicolor\_N/N\*n

> stopifnot(setosa\_n+virginica\_n+versicolor\_n == n)

> setosa\_sample\_i <- sample(setosa\_population\_i,setosa\_n,replace=FALSE)

> print(setosa\_sample\_i)

[1] 1 14 38 48 33 49 41 18 31 8

> virginica\_sample\_i <- sample(virginica\_population\_i,virginica\_n,FALSE)

> print(virginica\_sample\_i)

[1] 112 123 139 135 101 103 147 150 127 113

> versicolor\_sample\_i <- sample(versicolor\_population\_i,versicolor\_n,FALSE)

> print(versicolor\_sample\_i)

[1] 84 82 83 73 69 81 94 88 72 64

> setosa\_sample\_sepal\_length <- c(-1,setosa\_n)

> setosa\_sample\_sepal\_width <- c(-1,setosa\_n)

> setosa\_sample\_petal\_length <- c(-1,setosa\_n)

> setosa\_sample\_petal\_width <- c(-1,setosa\_n)

> for (i in 1:setosa\_n){

+ setosa\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[setosa\_sample\_i[i]]

+ setosa\_sample\_sepal\_width[i] <- pop .... [TRUNCATED]

> versicolor\_sample\_sepal\_length <- c(-1,versicolor\_n)

> versicolor\_sample\_sepal\_width <- c(-1,versicolor\_n)

> versicolor\_sample\_petal\_length <- c(-1,versicolor\_n)

> versicolor\_sample\_petal\_width <- c(-1,versicolor\_n)

> for (i in 1:versicolor\_n){

+ versicolor\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[versicolor\_sample\_i[i]]

+ versicolor\_sample\_sepal .... [TRUNCATED]

> virginica\_sample\_sepal\_length <- c(-1,virginica\_n)

> virginica\_sample\_sepal\_width <- c(-1,virginica\_n)

> virginica\_sample\_petal\_length <- c(-1,virginica\_n)

> virginica\_sample\_petal\_width <- c(-1,virginica\_n)

> for (i in 1:virginica\_n){

+ virginica\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[virginica\_sample\_i[i]]

+ virginica\_sample\_sepal\_wid .... [TRUNCATED]

> estimate\_sepal\_length\_mean = setosa\_N/N\*mean(setosa\_sample\_sepal\_length) + versicolor\_N/N\*mean(versicolor\_sample\_sepal\_length) + virginica\_N/N\*mean( .... [TRUNCATED]

> std\_err\_sepal\_length\_mean = sqrt((setosa\_N/N)^2\*(1-setosa\_n/setosa\_N)\*sd(setosa\_sample\_sepal\_length)^2/setosa\_n + (versicolor\_N/N)^2\*(1-versicolor\_n .... [TRUNCATED]

> sprintf("Estimate for population sepal length mean: %f",estimate\_sepal\_length\_mean)

[1] "Estimate for population sepal length mean: 5.763333"

> sprintf("Standard Error for estimate of population sepal length mean: %f",std\_err\_sepal\_length\_mean)

[1] "Standard Error for estimate of population sepal length mean: 0.072166"

> estimate\_sepal\_width\_mean = setosa\_N/N\*mean(setosa\_sample\_sepal\_width) + versicolor\_N/N\*mean(versicolor\_sample\_sepal\_width) + virginica\_N/N\*mean(vir .... [TRUNCATED]

> std\_err\_sepal\_width\_mean = sqrt((setosa\_N/N)^2\*(1-setosa\_n/setosa\_N)\*sd(setosa\_sample\_sepal\_width)^2/setosa\_n + (versicolor\_N/N)^2\*(1-versicolor\_n/v .... [TRUNCATED]

> sprintf("Estimate for population sepal width mean: %f",estimate\_sepal\_width\_mean)

[1] "Estimate for population sepal width mean: 2.950000"

> sprintf("Standard Error for estimate of population sepal width mean: %f",std\_err\_sepal\_width\_mean)

[1] "Standard Error for estimate of population sepal width mean: 0.043558"

> estimate\_petal\_length\_mean = setosa\_N/N\*mean(setosa\_sample\_petal\_length) + versicolor\_N/N\*mean(versicolor\_sample\_petal\_length) + virginica\_N/N\*mean( .... [TRUNCATED]

> std\_err\_petal\_length\_mean = sqrt((setosa\_N/N)^2\*(1-setosa\_n/setosa\_N)\*sd(setosa\_sample\_petal\_length)^2/setosa\_n + (versicolor\_N/N)^2\*(1-versicolor\_n .... [TRUNCATED]

> sprintf("Estimate for population petal length mean: %f",estimate\_petal\_length\_mean)

[1] "Estimate for population petal length mean: 3.703333"

> sprintf("Standard Error for estimate of population petal length mean: %f",std\_err\_petal\_length\_mean)

[1] "Standard Error for estimate of population petal length mean: 0.079944"

> estimate\_petal\_width\_mean = setosa\_N/N\*mean(setosa\_sample\_petal\_width) + versicolor\_N/N\*mean(versicolor\_sample\_petal\_width) + virginica\_N/N\*mean(vir .... [TRUNCATED]

> std\_err\_petal\_width\_mean = sqrt((setosa\_N/N)^2\*(1-setosa\_n/setosa\_N)\*sd(setosa\_sample\_petal\_width)^2/setosa\_n + (versicolor\_N/N)^2\*(1-versicolor\_n/v .... [TRUNCATED]

> sprintf("Estimate for population petal width mean: %f",estimate\_petal\_width\_mean)

[1] "Estimate for population petal width mean: 1.136667"

> sprintf("Standard Error for estimate of population petal width mean: %f",std\_err\_petal\_width\_mean)

[1] "Standard Error for estimate of population petal width mean: 0.034124"

> # Q3

> n = 30

> data\_i <- population\_data$X

> N = length(data\_i) # population size

> setosa\_osd = 0.35

> versicolor\_osd = 0.52

> virginica\_osd = 0.64

> setosa\_N = 0

> versicolor\_N = 0

> virginica\_N = 0

> for (i in 1:N){

+ type <- population\_data$Species[i]

+ if(type == 'setosa'){

+ setosa\_N = setosa\_N+1

+ }else if(type == 'versicolor'){

+ .... [TRUNCATED]

> stopifnot(setosa\_N+virginica\_N+versicolor\_N == N)

> setosa\_population\_i <- c(-1,setosa\_N)

> x <- 1

> versicolor\_population\_i <- c(-1,versicolor\_N)

> y <- 1

> virginica\_population\_i <- c(-1,virginica\_N)

> z <- 1

> for (i in 1:N){

+ type <- population\_data$Species[i]

+ if(type == 'setosa'){

+ setosa\_population\_i[x] = population\_data$X[i]

+ x = x+1

+ .... [TRUNCATED]

> setosa\_n = round(n\*(setosa\_N\*setosa\_osd)/(setosa\_N\*setosa\_osd+versicolor\_N\*versicolor\_osd+virginica\_N\*virginica\_osd))

> virginica\_n = round(n\*(virginica\_N\*virginica\_osd)/(setosa\_N\*setosa\_osd+versicolor\_N\*versicolor\_osd+virginica\_N\*virginica\_osd))

> versicolor\_n = round(n\*(versicolor\_N\*versicolor\_osd)/(setosa\_N\*setosa\_osd+versicolor\_N\*versicolor\_osd+virginica\_N\*virginica\_osd))

> stopifnot(setosa\_n+virginica\_n+versicolor\_n == n)

> setosa\_sample\_i <- sample(setosa\_population\_i,setosa\_n,replace=FALSE)

> print(setosa\_sample\_i)

[1] 39 36 22 42 28 21 24

> virginica\_sample\_i <- sample(virginica\_population\_i,virginica\_n,FALSE)

> print(virginica\_sample\_i)

[1] 103 131 141 126 130 133 101 116 128 149 150 122 102

> versicolor\_sample\_i <- sample(versicolor\_population\_i,versicolor\_n,FALSE)

> print(versicolor\_sample\_i)

[1] 77 54 63 74 90 69 89 70 73 62

> setosa\_sample\_sepal\_length <- c(-1,setosa\_n)

> for (i in 1:setosa\_n){

+ setosa\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[setosa\_sample\_i[i]]

+ }

> versicolor\_sample\_sepal\_length <- c(-1,versicolor\_n)

> for (i in 1:versicolor\_n){

+ versicolor\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[versicolor\_sample\_i[i]]

+ }

> virginica\_sample\_sepal\_length <- c(-1,virginica\_n)

> for (i in 1:virginica\_n){

+ virginica\_sample\_sepal\_length[i] <- population\_data$Sepal.Length[virginica\_sample\_i[i]]

+ }

> estimate\_sepal\_length\_mean = setosa\_N/N\*mean(setosa\_sample\_sepal\_length) + versicolor\_N/N\*mean(versicolor\_sample\_sepal\_length) + virginica\_N/N\*mean( .... [TRUNCATED]

> std\_err\_sepal\_length\_mean = sqrt((setosa\_N/N)^2\*(1-setosa\_n/setosa\_N)\*sd(setosa\_sample\_sepal\_length)^2/setosa\_n + (versicolor\_N/N)^2\*(1-versicolor\_n .... [TRUNCATED]

> sprintf("Estimate for population sepal length mean: %f",estimate\_sepal\_length\_mean)

[1] "Estimate for population sepal length mean: 5.797253"

> sprintf("Standard Error for estimate of population sepal length mean: %f",std\_err\_sepal\_length\_mean)

[1] "Standard Error for estimate of population sepal length mean: 0.074937"

>