Ram Yoogesh Gopu A3 - 20867060

Setting up

```
knitr::opts_chunk$set(echo = TRUE,
                      warning = FALSE,
                      message = FALSE,
                      fig.align = "center",
                      fig.width = 7,
                      fig.height = 6,
                      out.height = "60%")
set.seed(12314159)
library(loon.data)
library(loon)
```

```
## Loading required package: tcltk
```

```
library(gridExtra)

imageDirectory <- "./img"
dataDirectory <- "./data"
codeDirectory <- "/home/yoogesh/Desktop/stat847assignments/a3/"
path_concat <- function(path1, path2, sep="/") paste(path1, path2, sep = sep)


imageDirectory <- "/home/yoogesh/Desktop/stat847assignments/a3/img"  # e.g. in current "./img"
dataDirectory <- "/home/yoogesh/Desktop/stat847assignments/a3/data"  # e.g. in current "./data"
path_concat <- function(path1, path2, sep="/") paste(path1, path2, sep = sep)
```

(B)

```
load("blocks.rda")
head(blocks, n=3)
```

```
##   id weight perimeter group
## 1  1     55        32     B
## 2  2     35        27     B
## 3  3     35        25     A
```

(A)

(i)

```
set.seed(314159)
```

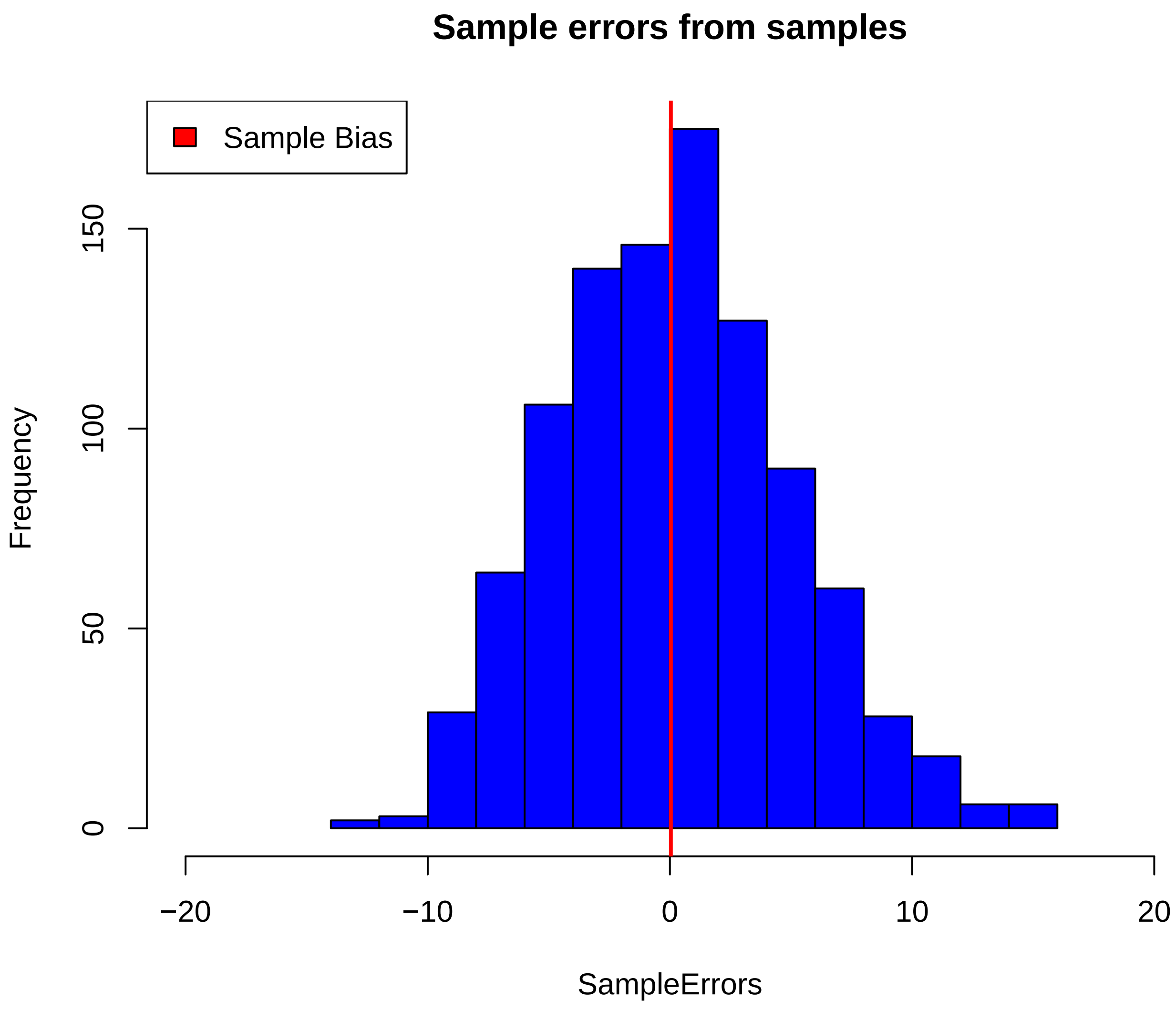
```r
randomsamp <- sapply(1:1000, function(x){

  return(list(sample(blocks$weight, size=10, replace = F)))
})

randomSampleAves <- sapply(randomsamp, mean)
```

(ii)

```r
## W defines weight for reference
avgW <- mean(blocks$weight)
sampleErrors <- sapply(randomSampleAves, function(x) {
  return((x - avgW))
})


sampleBias <- mean(sampleErrors)
avgSampW <- mean(randomSampleAves)
sampleVariability <- mean(sapply(randomSampleAves, function(x) {
  return((x - avgSampW) ^ 2)
}))

sampleMSE <- mean(sapply(randomSampleAves, function(x) {
  return((x - avgW) ^ 2)
}))
```

(iii)

```r
hist(sampleErrors, col = "blue", main = "Sample errors from samples", xlab = "SampleErrors", xlim = c(-2
abline(v = sampleBias, col="red", lwd=2)
legend("topleft", c("Sample Bias"), fill=c("red"))
```



##(B) ## (i)

```
set.seed(314159)

bone <- blocks$weight[blocks$group == 'A']
btwo <- blocks$weight[blocks$group == 'B']

stratSamples <- sapply(1:1000, function(x){
  return(list(append(sample(bone, size=5, replace = F), sample(btwo, size=5, replace = F))))
})

stratifiedSampleAves <- sapply(stratSamples, mean)
```

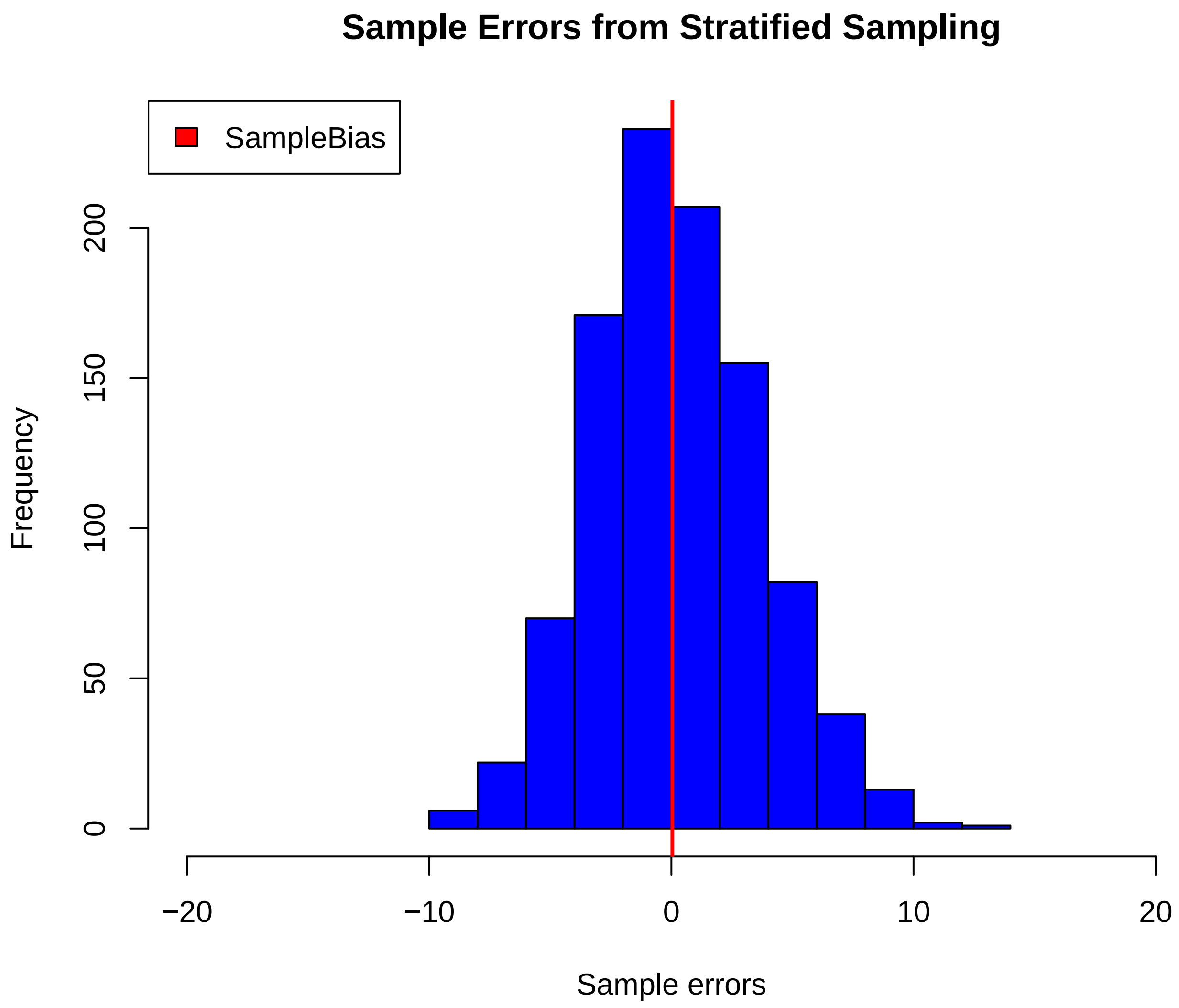
(ii)

```
set.seed(314159)

avgWStratS <- mean(blocks$weight)
```

```
sampleErrorsStrat <- sapply(stratifiedSampleAves, function(x)
{
  return(x - avgWStratS)
})

sampleBiasStrat <- mean(sampleErrorsStrat)
avgSampleWStratS <- mean(stratifiedSampleAves)

sampleVariabilityStrat <- mean(sapply(stratifiedSampleAves, function(x)
{
  return((x - avgSampleWStratS) ^ 2)
}))

sampleMSEStrat <- mean(sapply(stratifiedSampleAves, function(x)
{
  return((x - avgWStratS) ^ 2)
}))
```

(iii)

```
hist(sampleErrorsStrat, col = "Blue", main = "Sample Errors from Stratified Sampling", xlab = "Sample e
abline(v = sampleBias, col="red", lwd=2)
legend("topleft", c("SampleBias"), fill=c("red"))
```



(C)

(i)

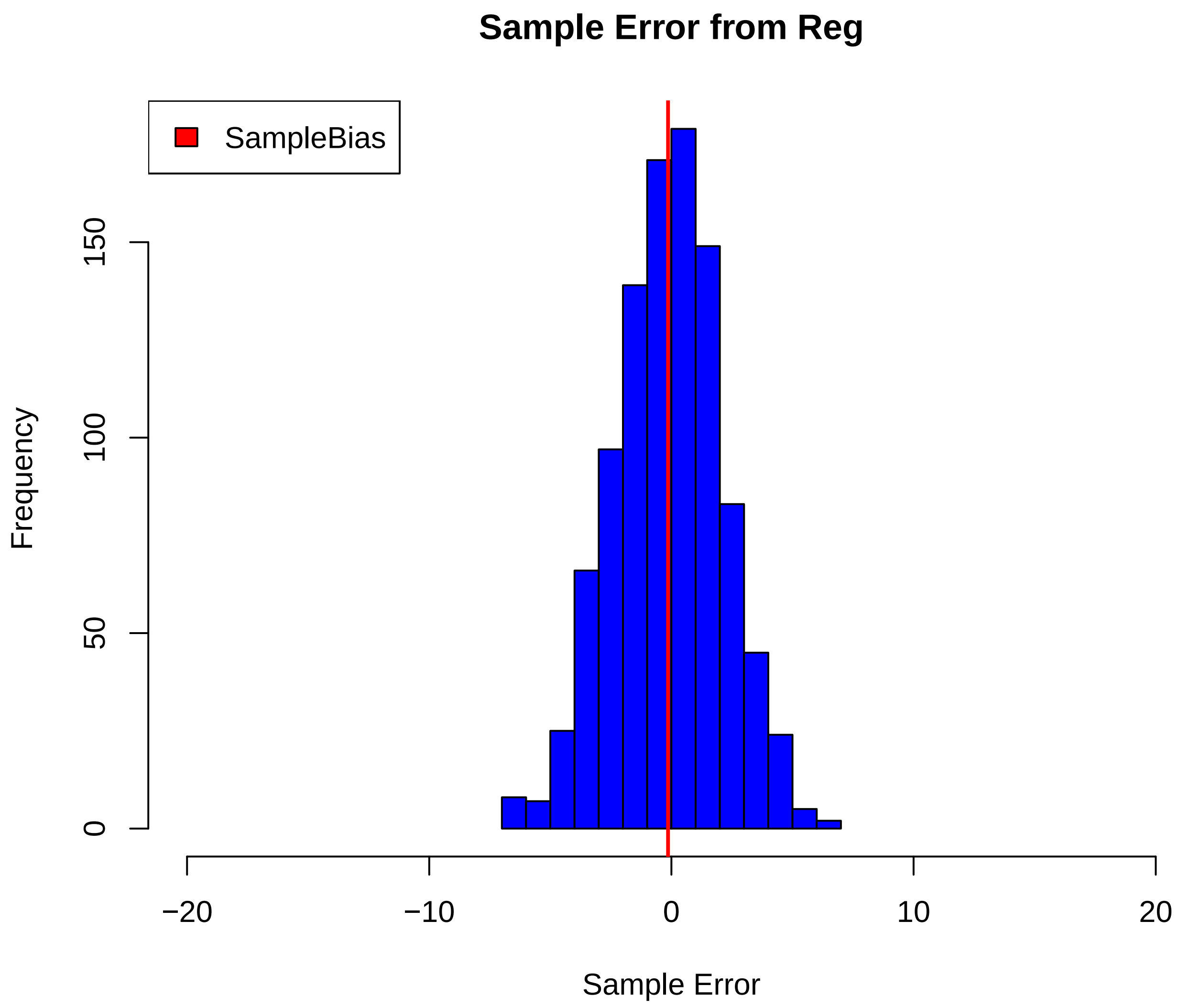
```r
set.seed(314159)

avgPeri <- mean(blocks$perimeter)
regressionEstimates <- sapply(1:1000, function(x)
{

  getid <- sample(1:100, size=10, replace = F)
  rmod <- lm(weight ~ perimeter, data=blocks[getid,])
  rdata = data.frame(perimeter=avgPeri)
  return(predict(rmod, rdata))

  })
```

##(ii)

```r
avgW <- mean(blocks$weight)

serrorsreg <- sapply(regressionEstimates, function(x)
{
  return(x - avgW)
})

sbiasreg <- mean(serrorsreg)
avgSWreg <- mean(regressionEstimates)

svarreg <- mean(sapply(regressionEstimates, function(x)
{
  return((x - avgSWreg) ^ 2)
}))

smsereg <- mean(sapply(regressionEstimates, function(x)
{
  return((x - avgW) ^ 2)
}))
```

(iii)

```r
hist(serrorsreg, col = "blue", main = "Sample Error from Reg", xlab = "Sample Error", xlim = c(-20,20))
abline(v = sbiasreg, col="red", lwd=2)
legend("topleft", c("SampleBias"), fill=c("red"))
```



(iv)

I would say it is true .

(D)

```
students <- read.csv("judgmentSamples.csv")
head(students, n = 3)
```

```
##   studentID first second third fourth fifth sixth seventh eighth ninth
## 1      5086    12     18    17     11    15    20      14     13    16
## 2      3848    34     35    70     56    32    14       5     88    81
## 3      6656    14     34    41     29    32    55      74     40    16
##   tenth
## 1    18
## 2    73
## 3    70
```

(i)

```
set.seed(314159)
judgmentAves <- apply(students, 1, function(s)
{
    mean(sapply(s[c(2:length(s))], function(x) {
    blocks$weight[blocks$id == x][1]

}))
})
print("Average")
```

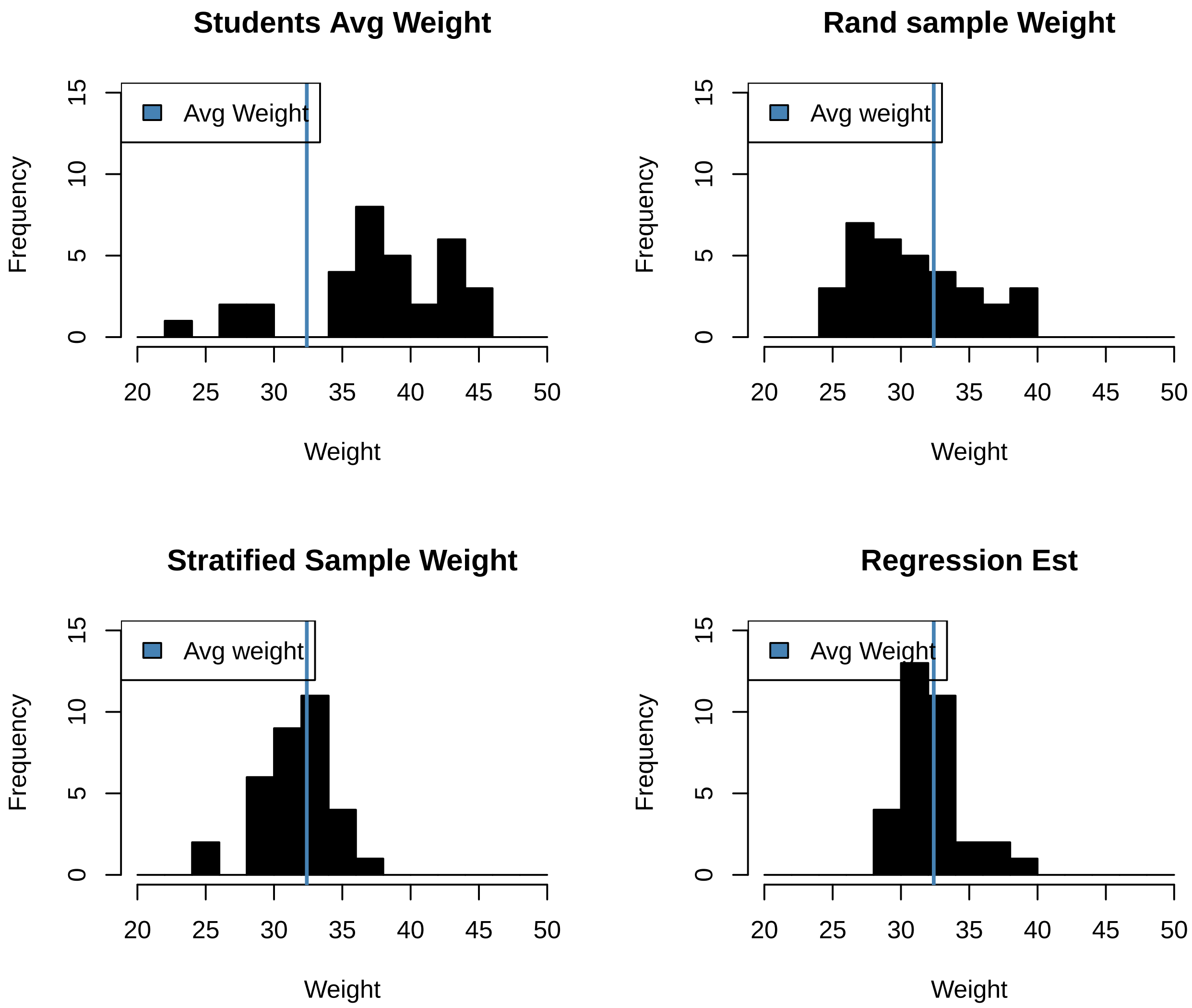
```
## [1] "Average"
```

```
mean(judgmentAves)
```

```
## [1] 37.81818
```

(ii)

```
set.seed(314159)

par(mfrow=c(2,2))

hist(judgmentAves,col="black",main = "Students Avg Weight",xlab = "Weight",xlim = c(20, 50),ylim = c(0,
abline(v = avgW, col="steelblue", lwd=2)
legend("topleft", c("Avg Weight"), fill=c("steelblue"))


hist(randomSampleAves[1:nrow(students)],col="black",main = "Rand sample Weight",xlab = "Weight",xlim =
abline(v = avgW, col="steelblue", lwd=2)
legend("topleft", c("Avg weight"), fill=c("steelblue"))


hist(stratifiedSampleAves[1:nrow(students)],col ="black",main ="Stratified Sample Weight",xlab = "Weigh
     ylim = c(0,15), breaks = seq(20, 50, 2))
abline(v = avgW, col="steelblue", lwd=2)
legend("topleft", c("Avg weight"), fill=c("steelblue"))


hist(regressionEstimates[1:nrow(students)],col ="black",main = "Regression Est",xlab = "Weight",xlim =
abline(v = avgW, col="steelblue", lwd=2)
legend("topleft", c("Avg Weight"), fill=c("steelblue"))
```


Students Avg Weight
Avg Weight
Frequency
0
5
10
15
20
25
30
35
40
45
50
Weight
Rand sample Weight
Avg weight
Frequency
0
5
10
15
20
25
30
35
40
45
50
Weight
Stratified Sample Weight
Avg weight
Frequency
0
5
10
15
20
25
30
35
40
45
50
Weight
Regression Est
Avg Weight
Frequency
0
5
10
15
20
25
30
35
40
45
50
Weight

(E)