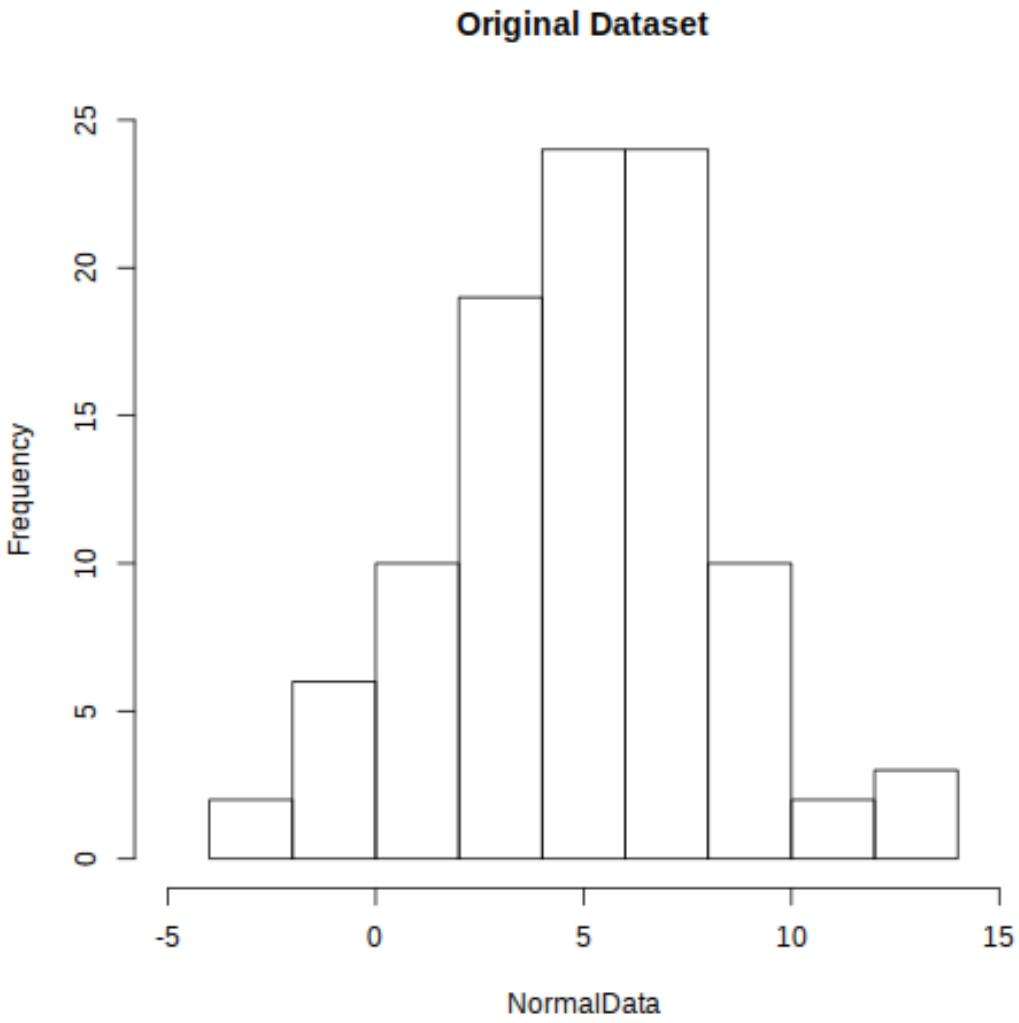
STA 032 R Homework 4

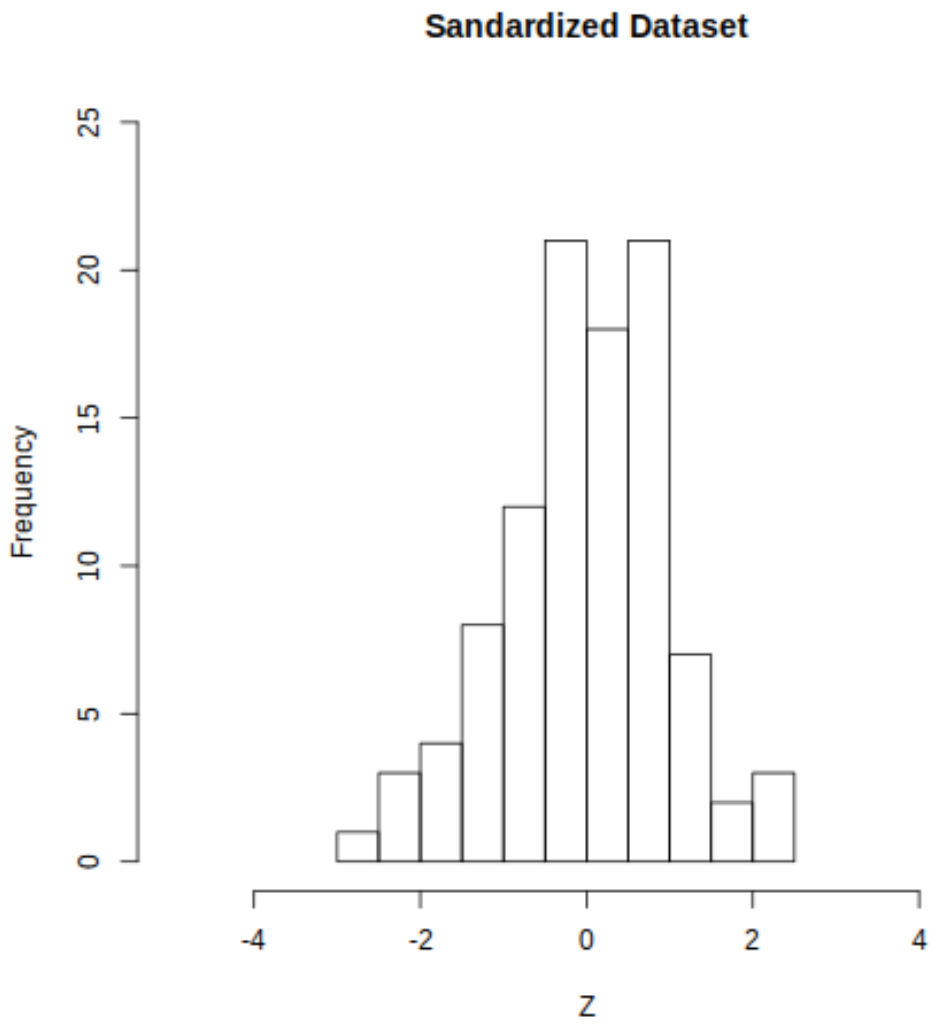
Hardy Jones
999397426
Professor Melcon
Winter 2015



1. (a)

   The mean of `NormalData` is 5.00.

   The standard deviation of `NormalData` is 3.35.

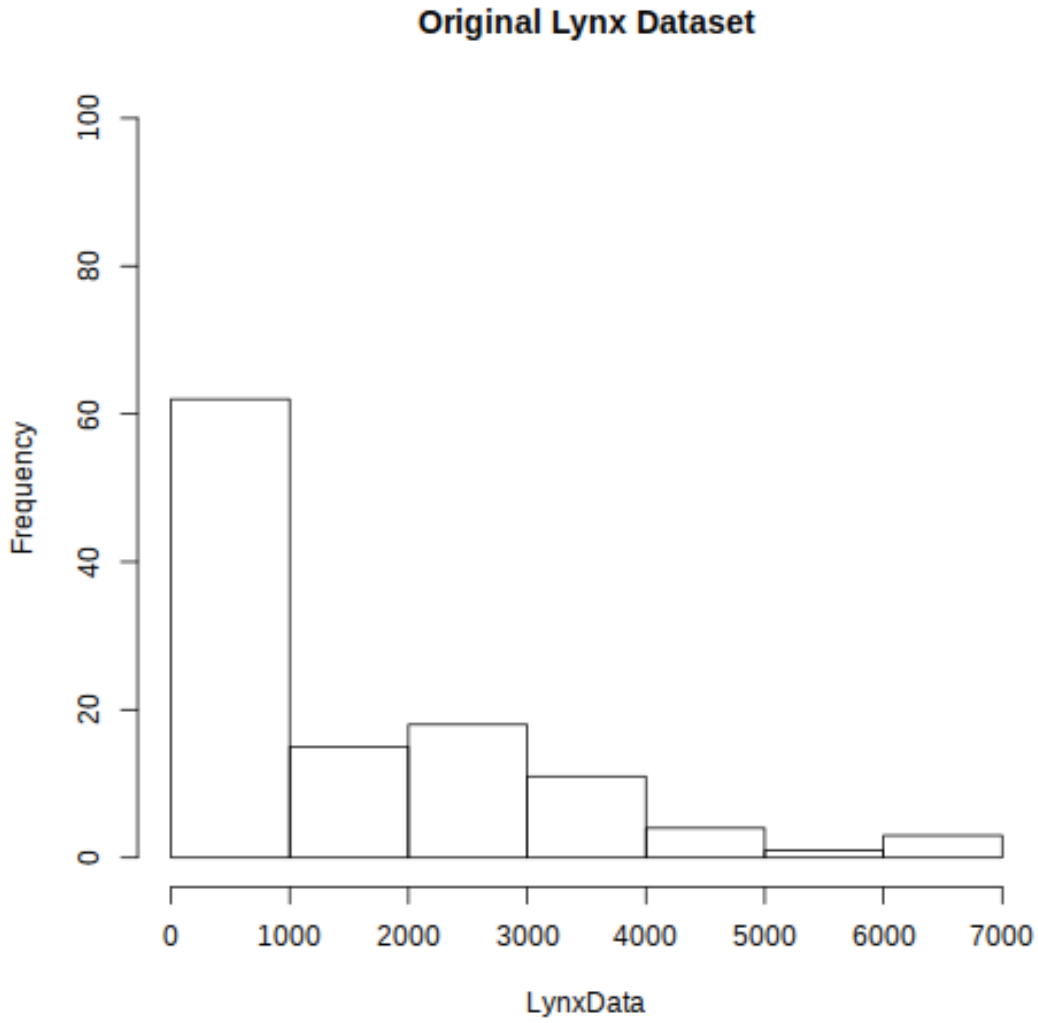


(b)

The mean of `Z` is -6.17×10^{-17}.

The standard deviation of `Z` is 1.

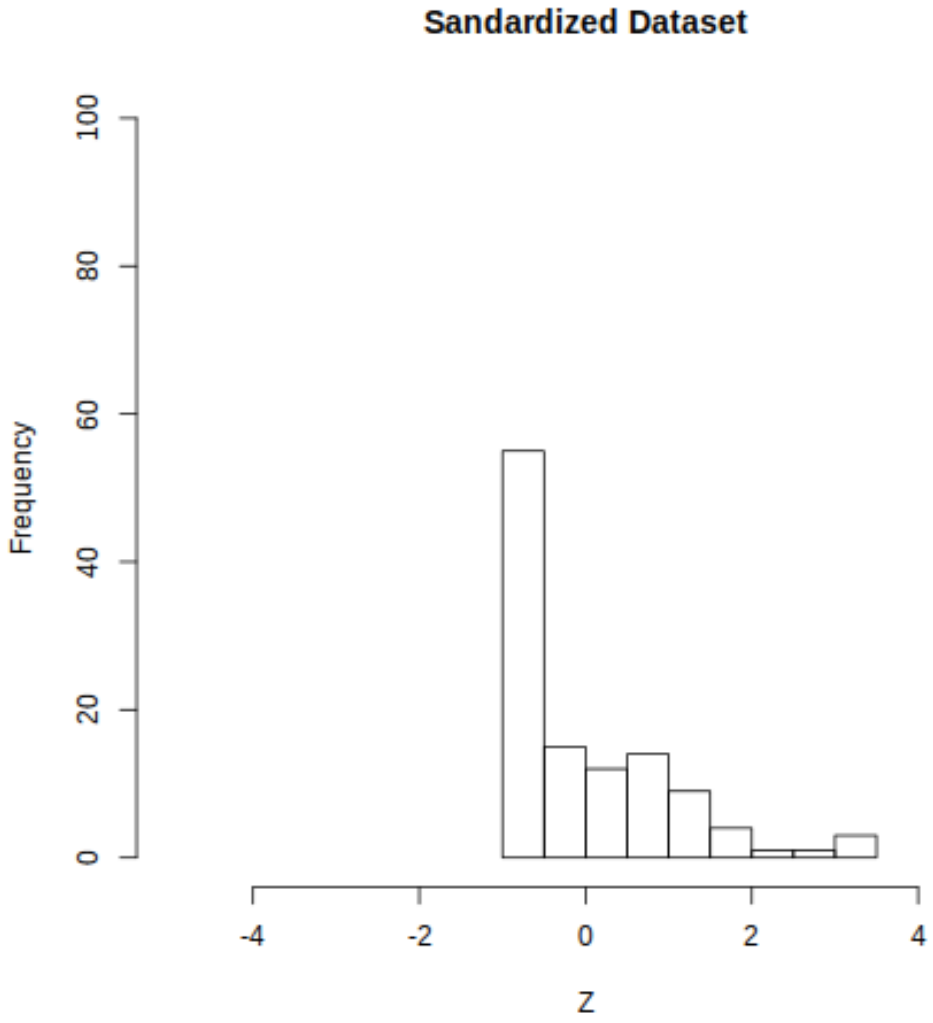
(c) The transformation moved the data so that it is now centered at 0. The data is also more heavily weighted towards 0. The range of data is also smaller. The height appears to be just a touch smaller than before. The overall shape is still the same, it still appears to be normal.



2. (a)

The mean of LynxData is 1.54×10^3.

The sd of LynxData is 1.59×10^3.

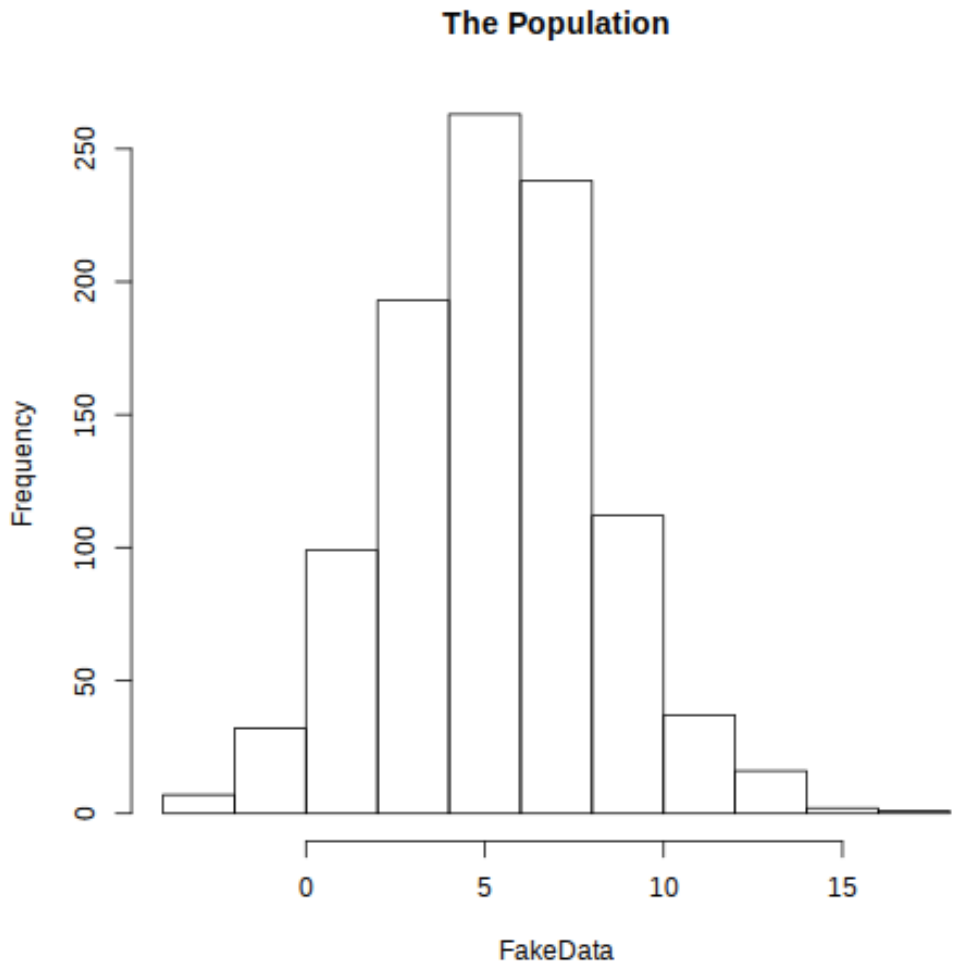


(b)

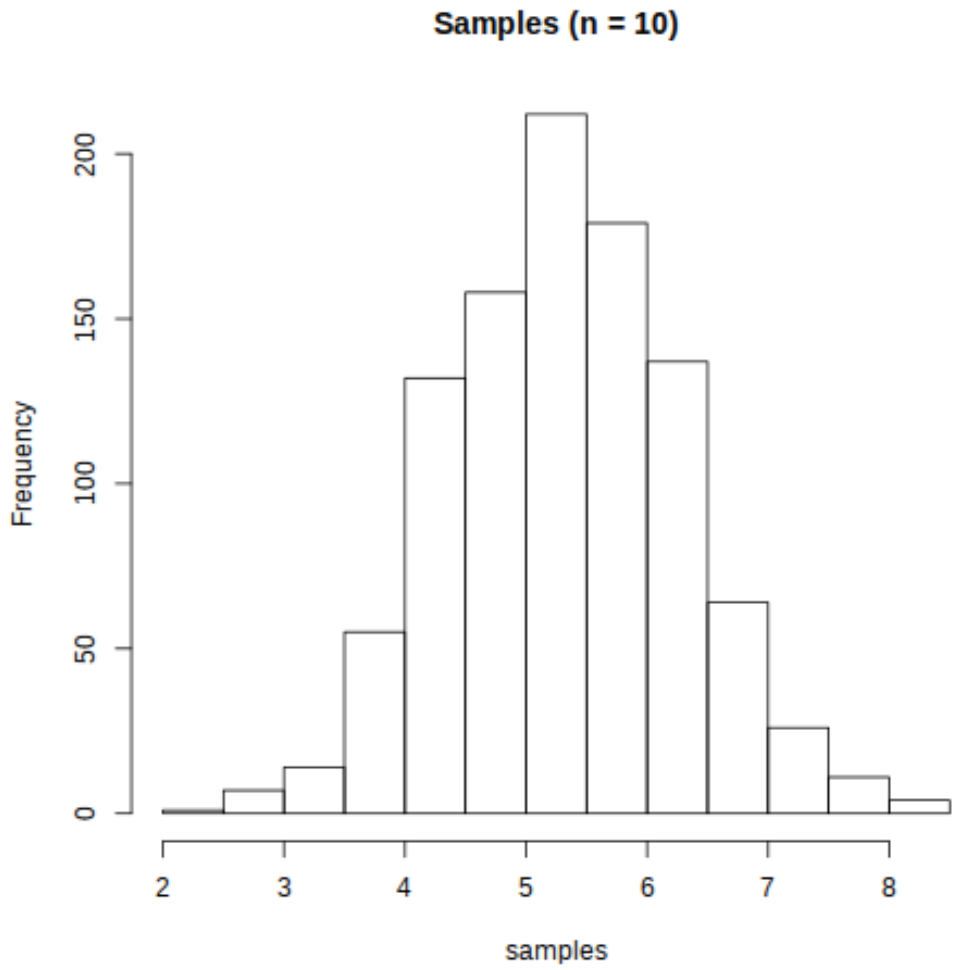
The mean of Z is 4.22×10^{-17}.

The sd of Z is 1.

(c) The transformation moved the data so that it is now much closer to 0. The data is also more heavily weighted towards 0. The range of data is also smaller. The height appears to be just a touch smaller than before. The overall shape is still the same.


The Population
Frequency
0
50
100
150
200
250
0
5
10
15
FakeData

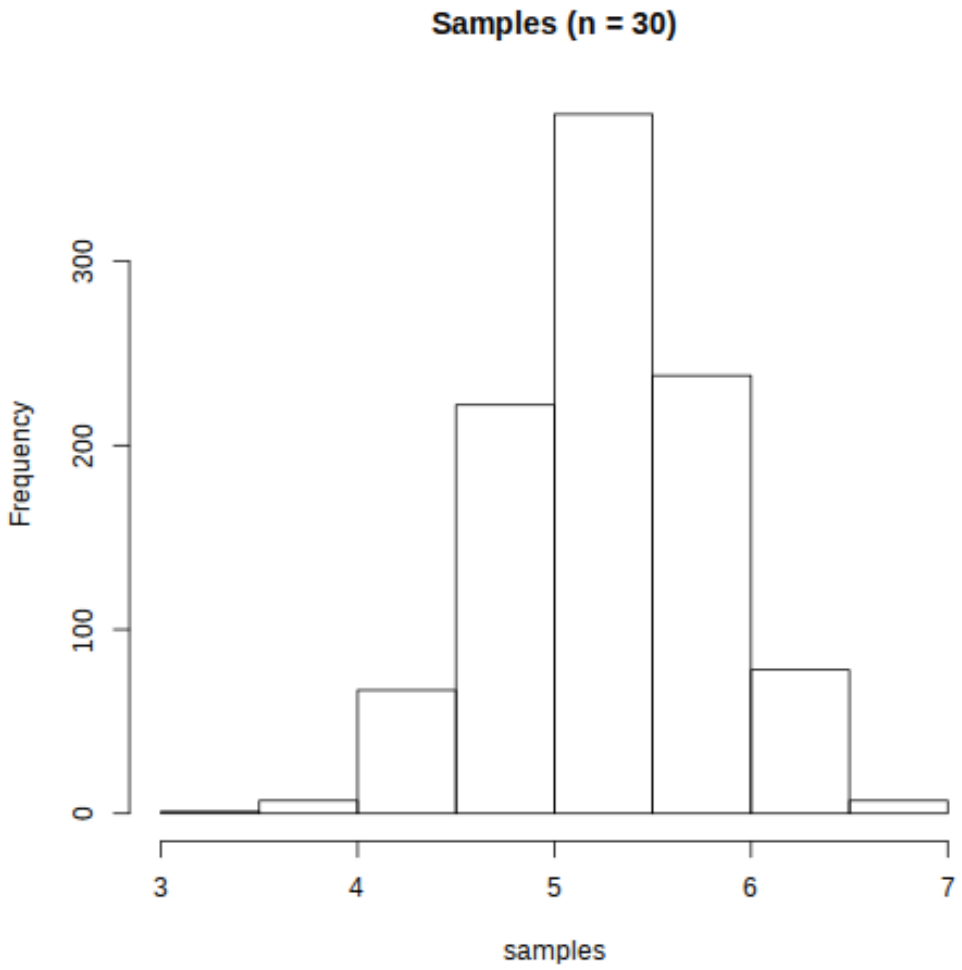
3. (a) i.



ii.

The mean for part ii is 4.80.

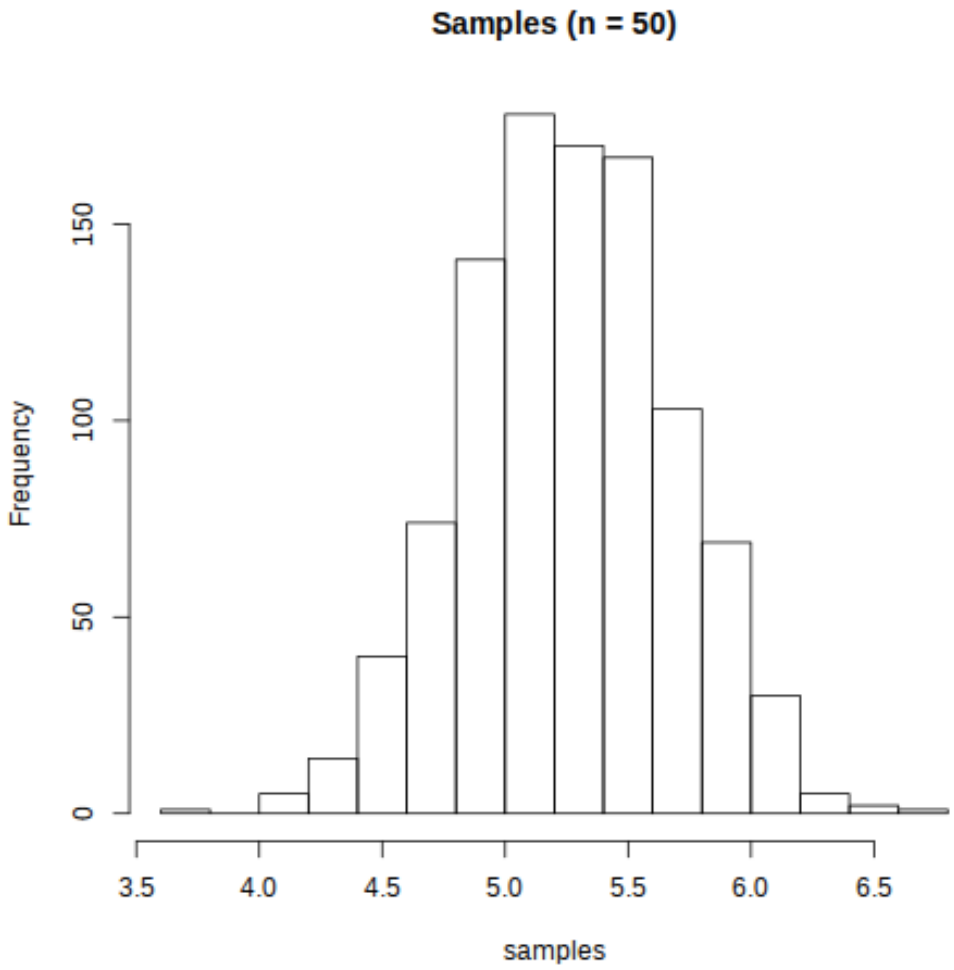
The sd for part ii is 9.31×10^{-1}.



iii.

The mean for part iii is 4.79.

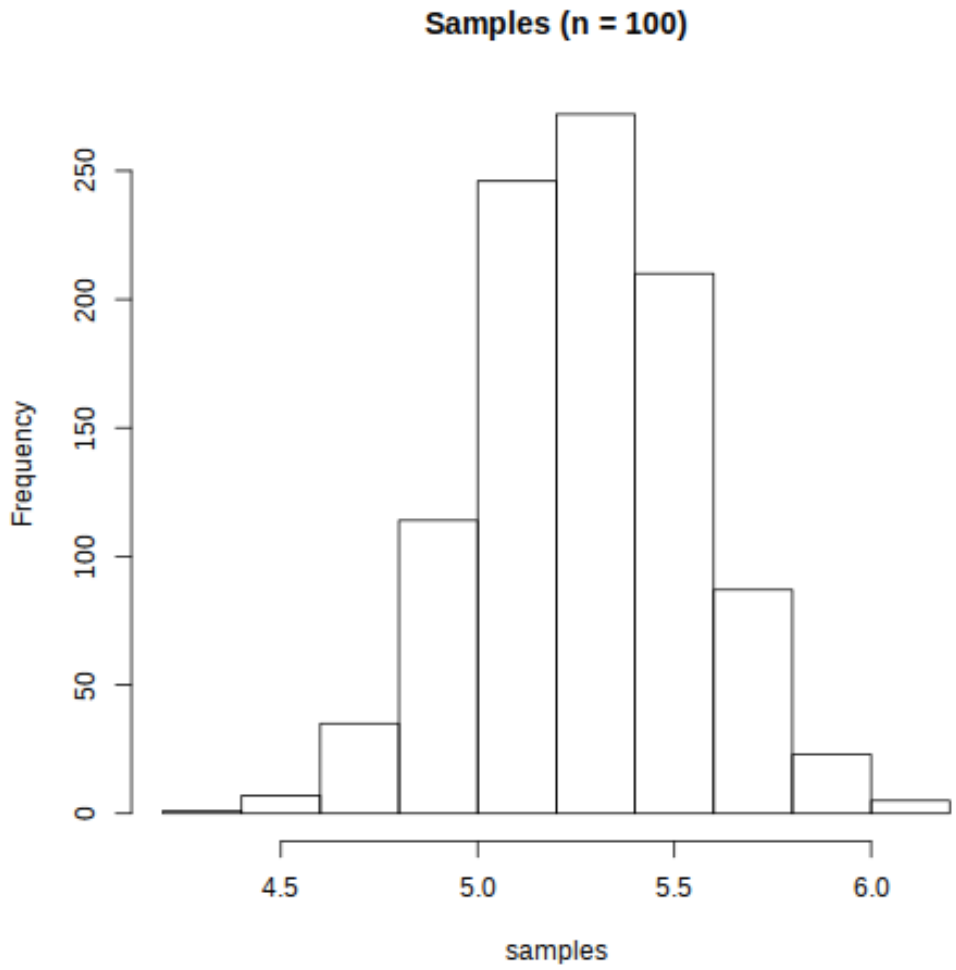
The sd for part iii is 5.52×10^{-1}.



iv.

The mean for part iv is 4.78.

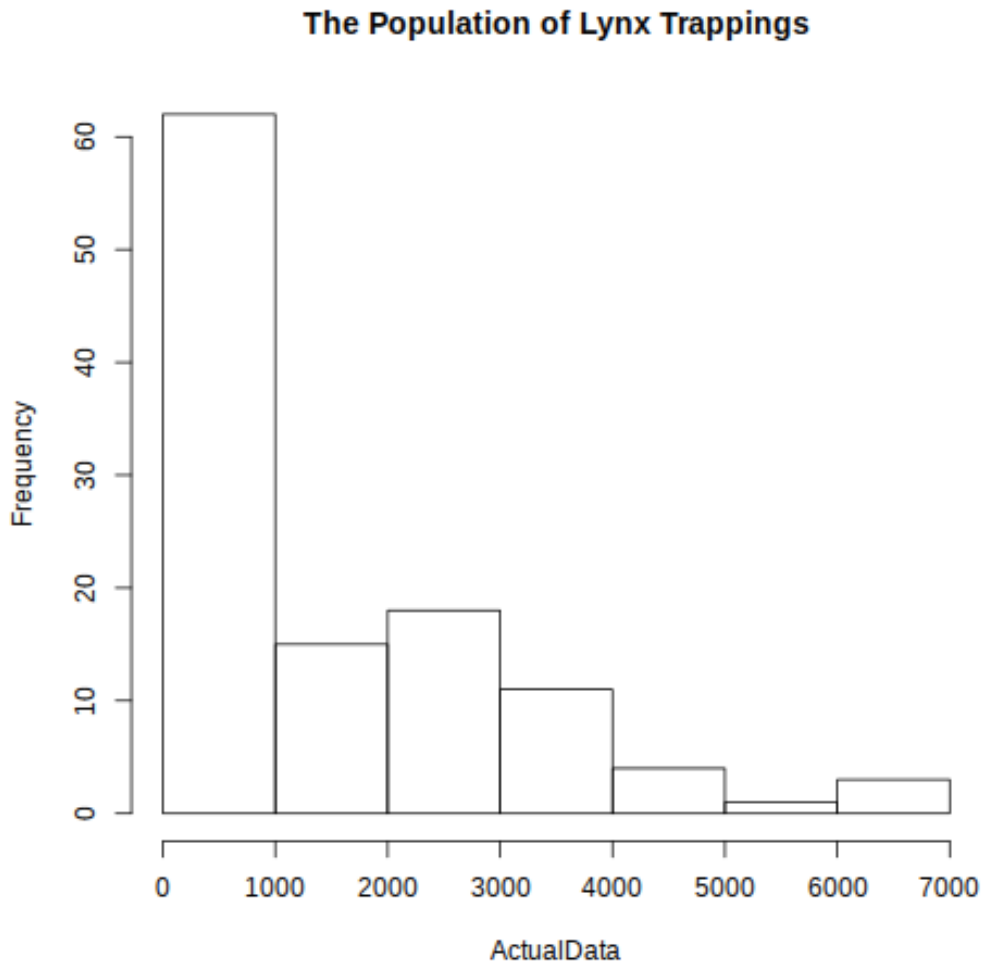
The sd for part iv is 4.37×10^{-1}.



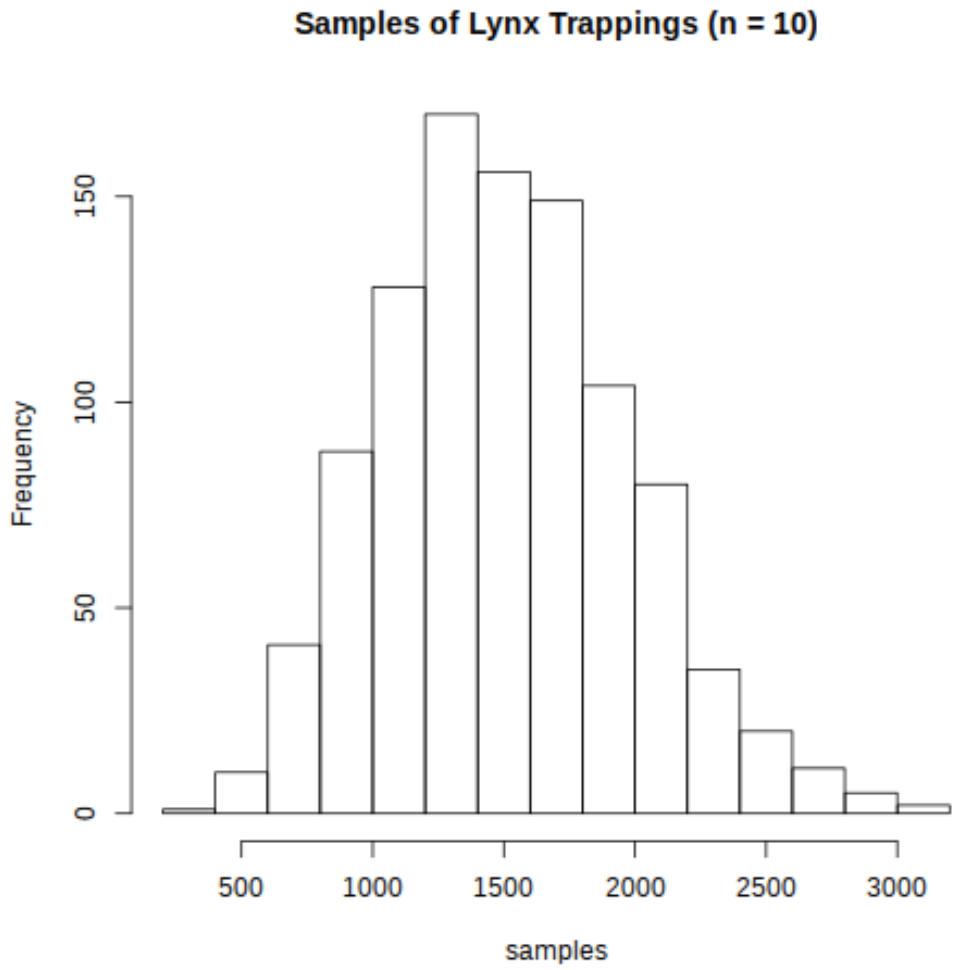
v.

The mean for part v is 4.77.

The sd for part v is 2.98×10^{-1}.

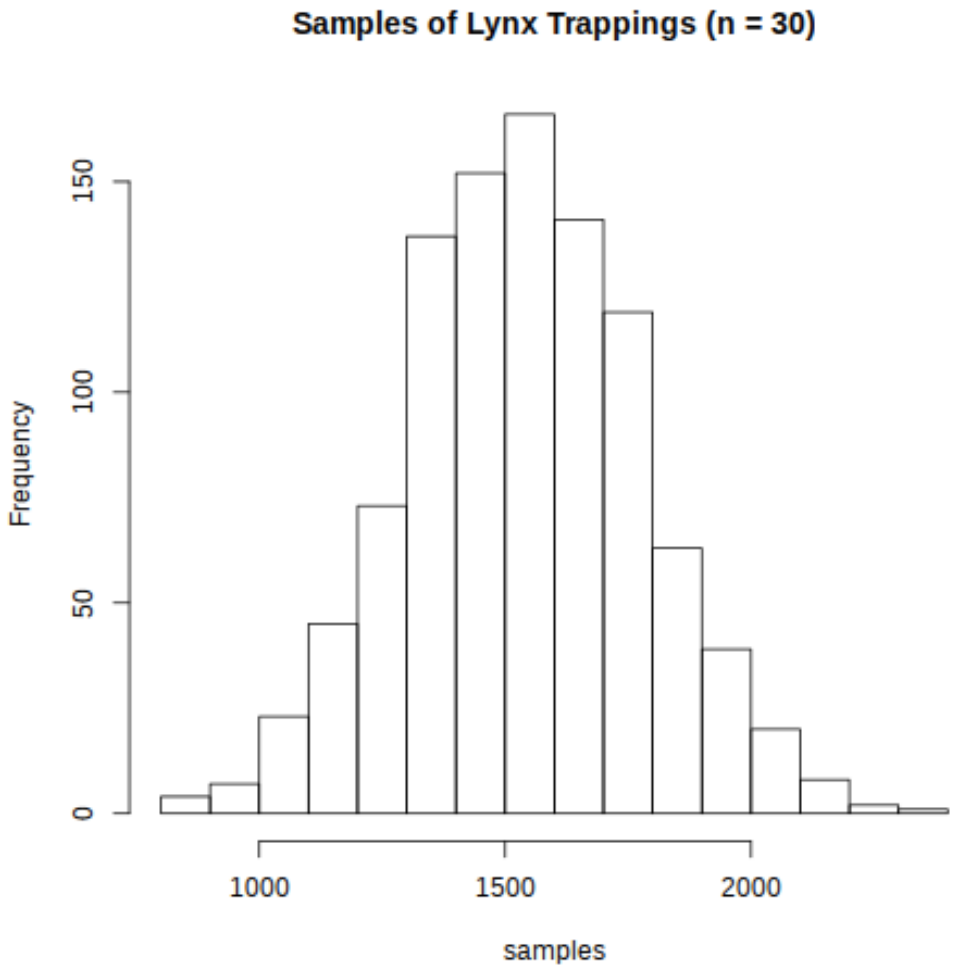

The Population of Lynx Trappings
Frequency
0
10
20
30
40
50
60
0
1000
2000
3000
4000
5000
6000
7000
ActualData

(b) i.



ii.

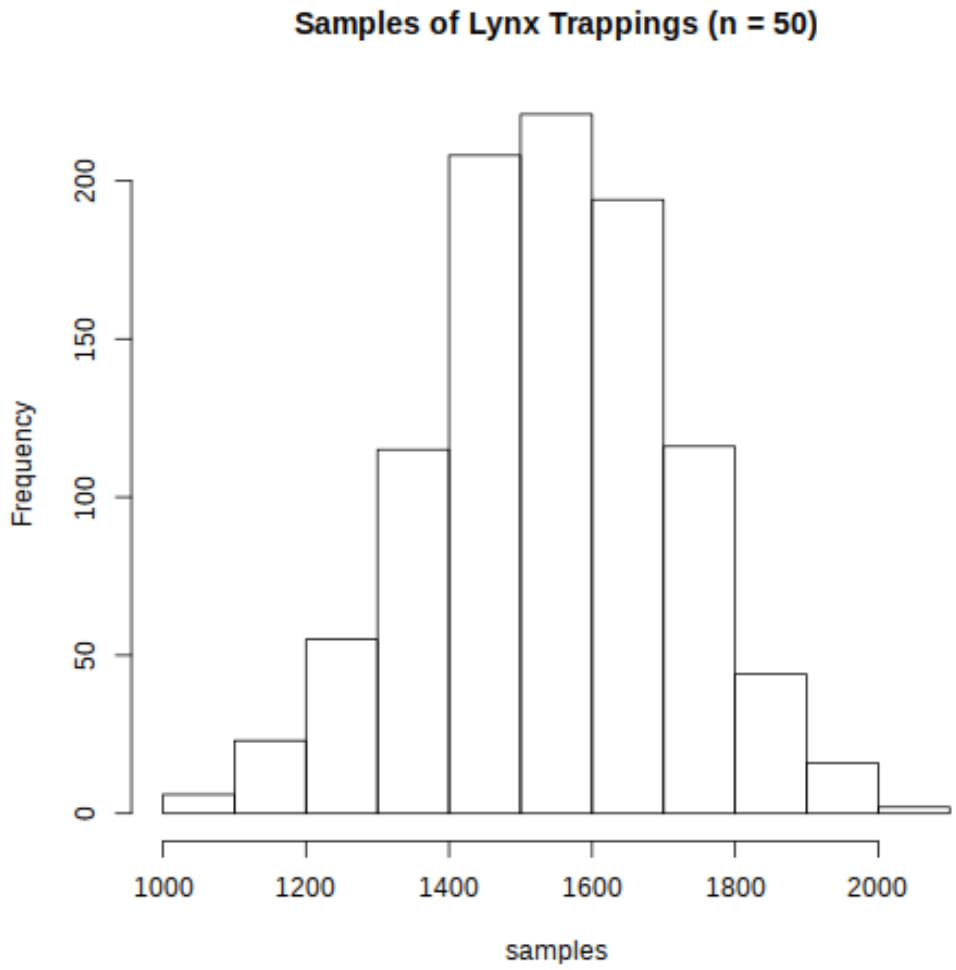
The mean for part ii is 1.52×10^3.
The sd for part ii is 4.62×10^2.



iii.

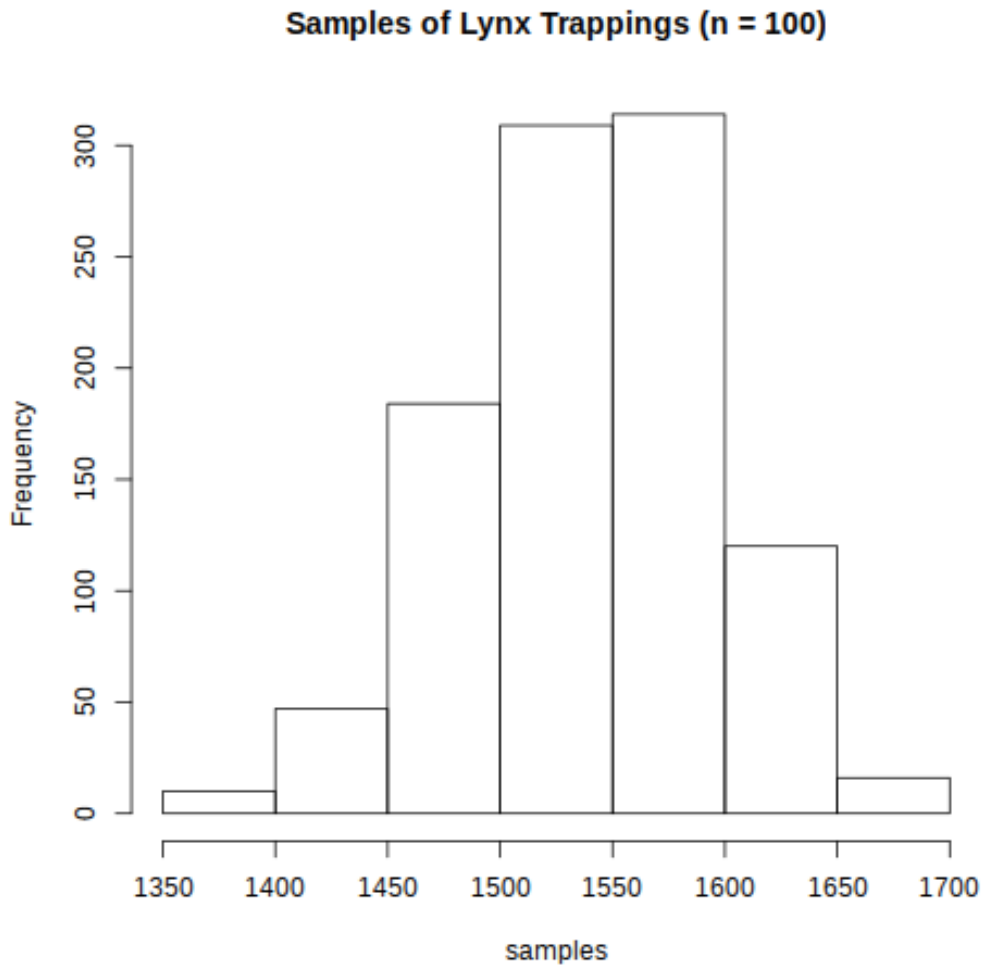
The mean for part iii is 1.53×10^3.

The sd for part iii is 2.52×10^2.



iv.

The mean for part iv is 1.54×10^3.
The sd for part iv is 1.64×10^2.



v.

The mean for part v is 1.54×10^3.
The sd for part v is 5.53×10^1.

(c) The distributions in parts a and b are both approximately normal distributions. It appears that as n increases, the range of values decreases.

Appendix A R code

Problem 1

```
NormalData <- rnorm(100, mean = 5, sd = 3)
```

(a)

```
png("prob1a.png")
hist(NormalData, main = "Original Dataset", xlim = c(-5, 15), ylim = c(0, 25))
dev.off()

mean(NormalData)

sd(NormalData)
```

(b)

```
Z <- (NormalData - mean(NormalData)) / sd(NormalData)

png("prob1b.png")
hist(Z, main = "Sandardized Dataset", xlim = c(-5, 5), ylim = c(0, 25))
dev.off()

mean(Z)

sd(Z)
```

Problem 2

```
LynxData <- as.numeric(lynx)
```

(a)

```
png("prob2a.png")
hist(LynxData, main = "Original Lynx Dataset", xlim = c(0, 7000), ylim = c(0, 100))
dev.off()

mean(LynxData)

sd(LynxData)
```

(b)

```
Z <- (LynxData - mean(LynxData)) / sd(LynxData)

png("prob2b.png")
hist(Z, main = "Sandardized Dataset", xlim = c(-5, 5), ylim = c(0, 100))
```

```
dev.off()

mean(Z)

sd(Z)
```

Problem 3

```
SampleMean <- function(X, n) {
  mean(sample(X, n))
}

SampleN <- function(X, n, N) {
  sapply(c(1:N), function(N) { SampleMean(X, n) })
}

SampleResults <- function(X, n, N, filename = "", main = "Samples") {
  samples <- SampleN(X, n, N)

  png(filename)
  hist(samples, main = main)
  dev.off()

  list(mean = mean(samples), sd = sd(samples))
}
```

(a)

```
source("prob3.R")

FakeData <- rnorm(1000, mean=5, sd=3)

png("prob3a.png")
hist(FakeData, main = "The Population")
dev.off()

ii  <- SampleResults(FakeData, 10, 1000,
    filename = "prob3a_ii.png", main = "Samples (n = 10)"
)
iii <- SampleResults(FakeData, 30, 1000,
    filename = "prob3a_iii.png", main = "Samples (n = 30)"
)
iv  <- SampleResults(FakeData, 50, 1000,
    filename = "prob3a_iv.png", main = "Samples (n = 50)"
)
v   <- SampleResults(FakeData, 100, 1000,
    filename = "prob3a_v.png", main = "Samples (n = 100)"
)

ii$mean
ii$sd
```

```
iii$mean
iii$sd

iv$mean
iv$sd

v$mean
v$sd
```

(b)

```
source("prob3.R")

ActualData <- as.numeric(lynx)

png("prob3b.png")
hist(ActualData, main = "The Population of Lynx Trappings")
dev.off()

ii  <- SampleResults(ActualData, 10, 1000,
    filename = "prob3b_ii.png", main = "Samples of Lynx Trappings (n = 10)"
)
iii <- SampleResults(ActualData, 30, 1000,
    filename = "prob3b_iii.png", main = "Samples of Lynx Trappings (n = 30)"
)
iv  <- SampleResults(ActualData, 50, 1000,
    filename = "prob3b_iv.png", main = "Samples of Lynx Trappings (n = 50)"
)
v   <- SampleResults(ActualData, 100, 1000,
    filename = "prob3b_v.png", main = "Samples of Lynx Trappings (n = 100)"
)

ii$mean
ii$sd

iii$mean
iii$sd

iv$mean
iv$sd

v$mean
v$sd
```