BRSM_ASSIGNMENT_1

Mafruz Nissar Rahman

2026-01-29

set environment

```r
# Set rendering parameters
knitr::opts_chunk$set(message = FALSE, warning = FALSE)



library(tidyverse) # data manipulation and visualization
library(tinytex)
```

Question 1

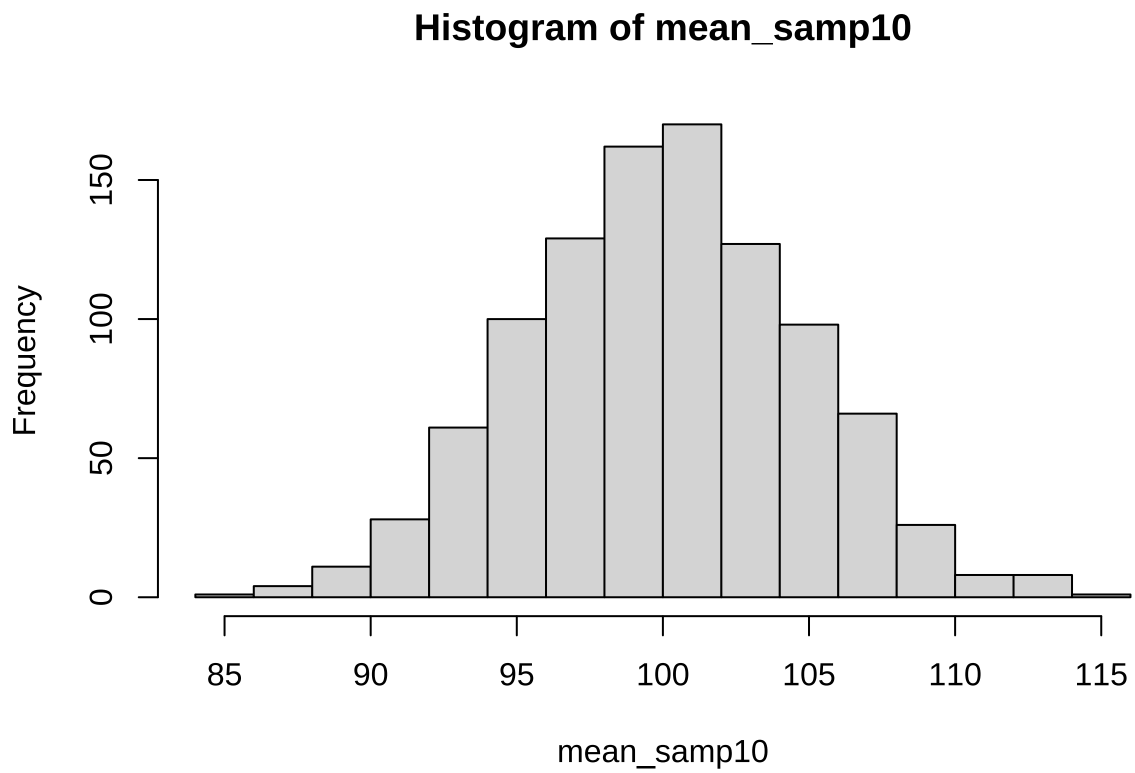
```r
samp1 = rnorm(10,100,15)
print(mean(samp1))
```

```
## [1] 94.40882
```

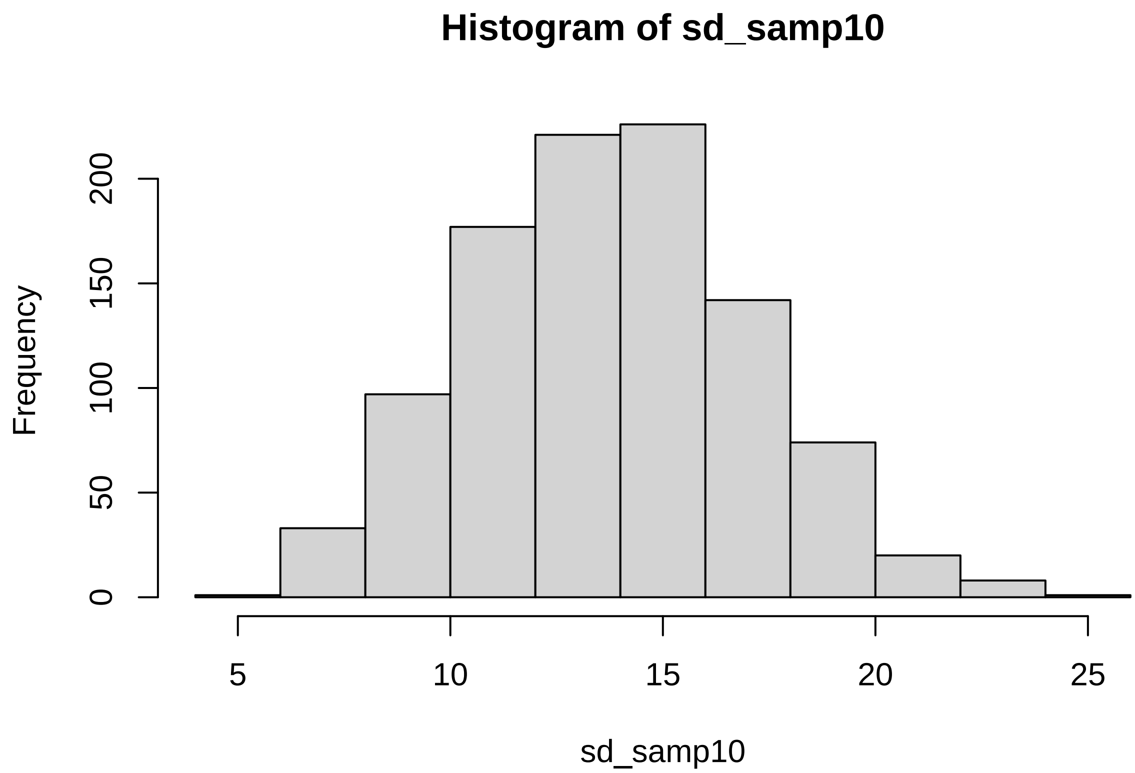
```r
var_samp1 = sum((samp1-mean(samp1))^2) / length(samp1)
sd_samp1 = sqrt(var_samp1)
print(sd_samp1)
```

```
## [1] 16.50669
```

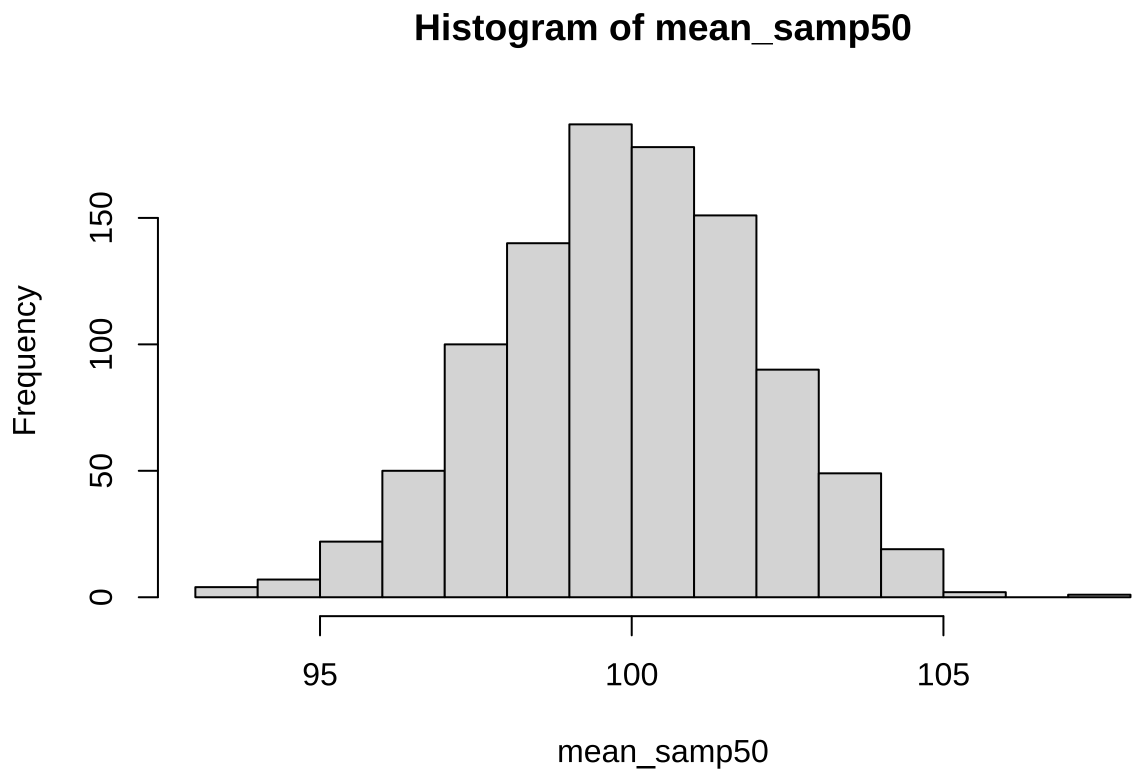
```r
# Sample Size = 10

mean_samp10 = numeric(1000)
sd_samp10 = numeric(1000)
for (ii in 1:1000) {
  samp10= rnorm(10,100,15)
  mean_samp10[ii] = mean(samp10)
  sd_samp10[ii] = sqrt(sum((samp10-mean(samp10))^2)/ length(samp10))
}

hist(mean_samp10)
```



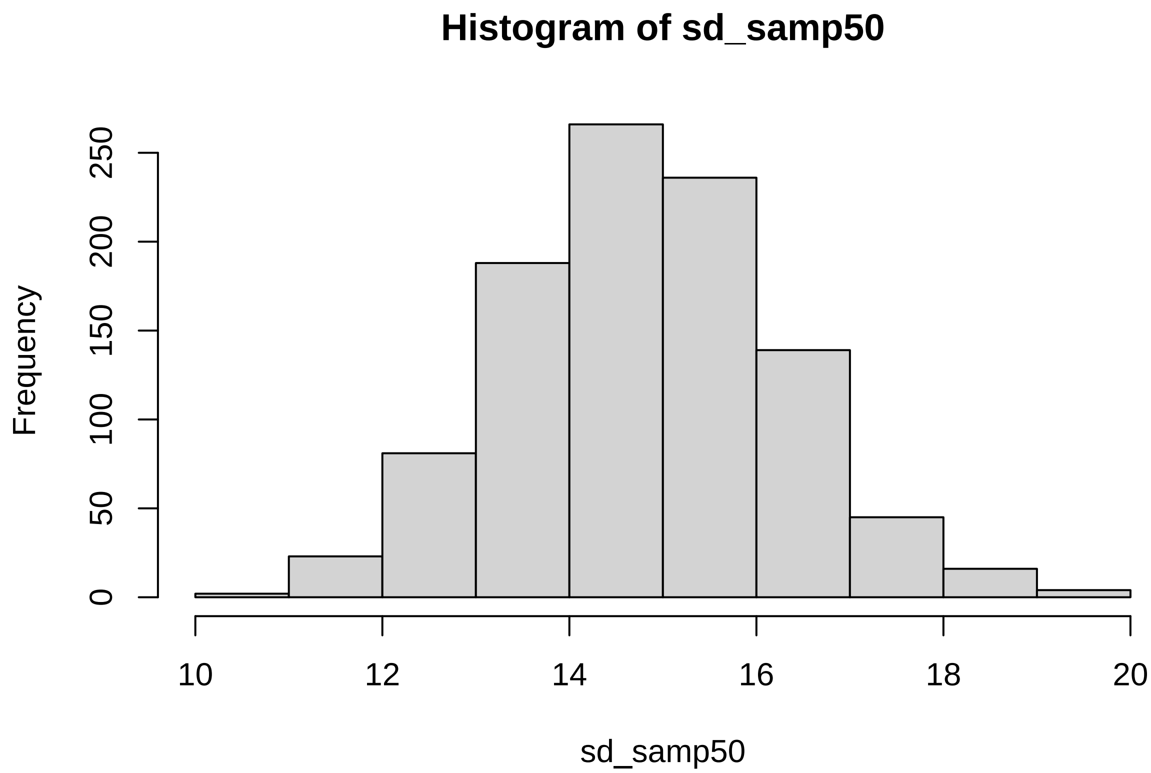

```
hist(sd_samp10)
```



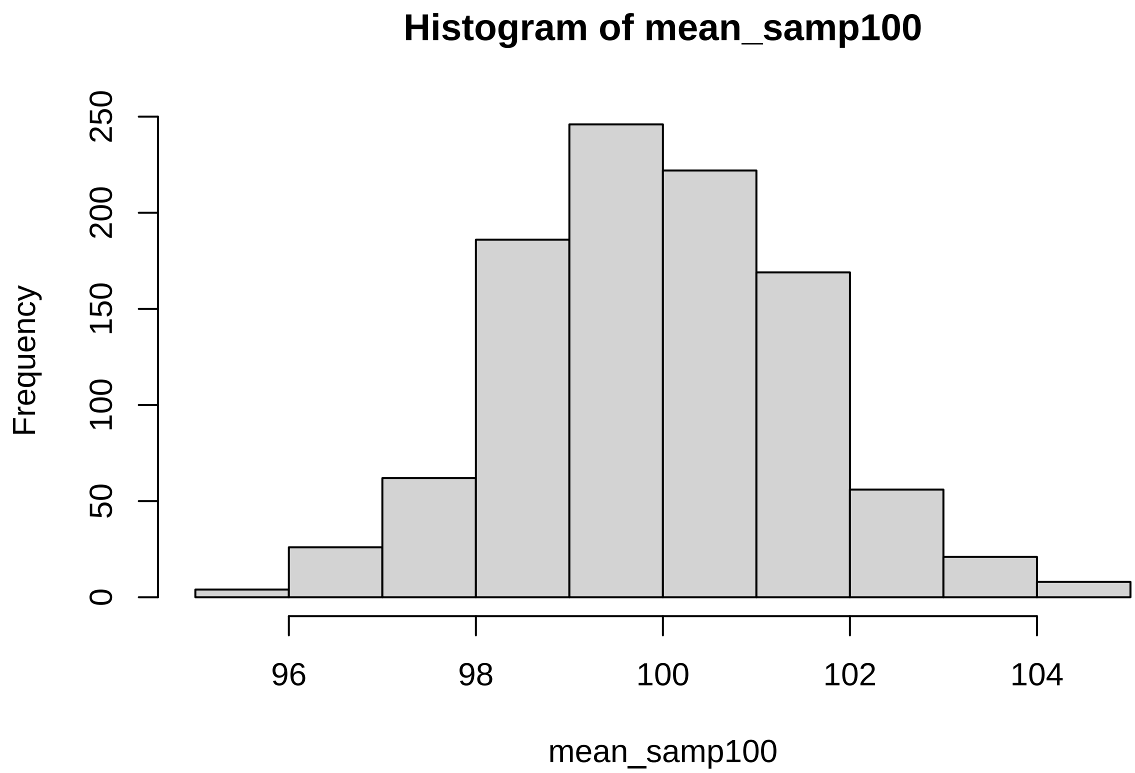

```r
# Sample Size = 50
mean_samp50 = numeric(1000)
sd_samp50 = numeric(1000)
for (ii in 1:1000) {
  samp50= rnorm(50,100,15)
  mean_samp50[ii] = mean(samp50)
  sd_samp50[ii] = sqrt(sum((samp50-mean(samp50))^2)/ length(samp50))
}

hist(mean_samp50)
```



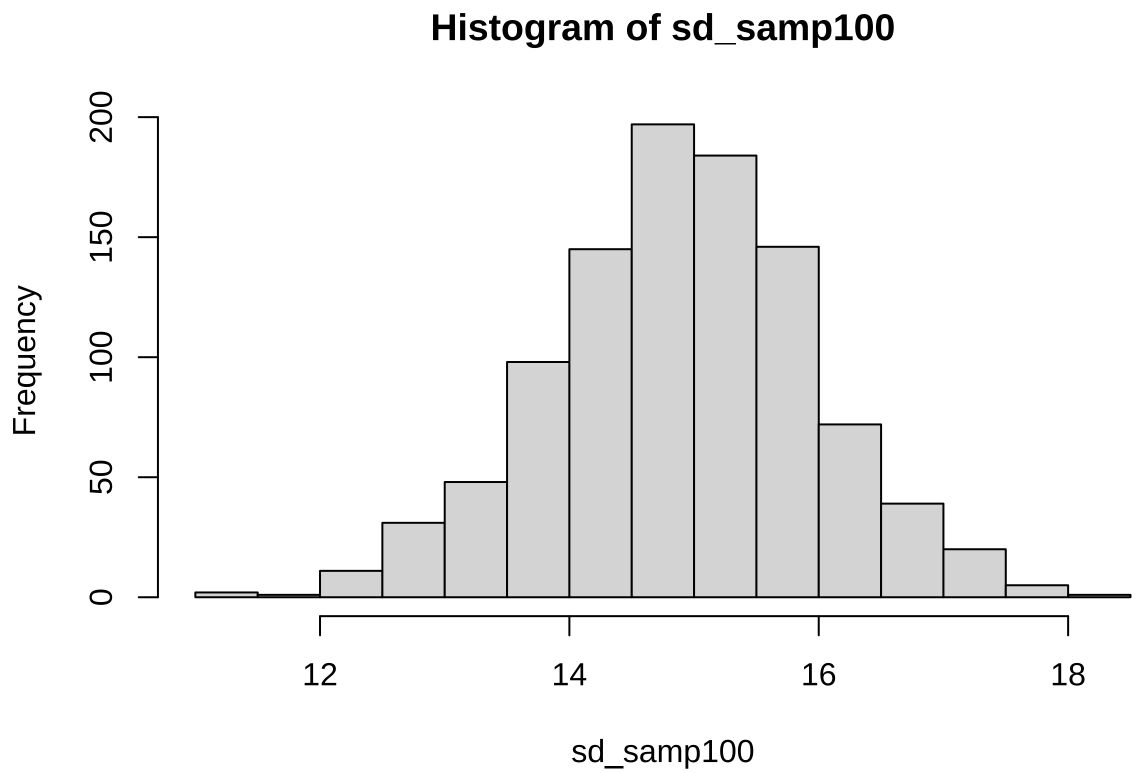

```r
hist(sd_samp50)
```



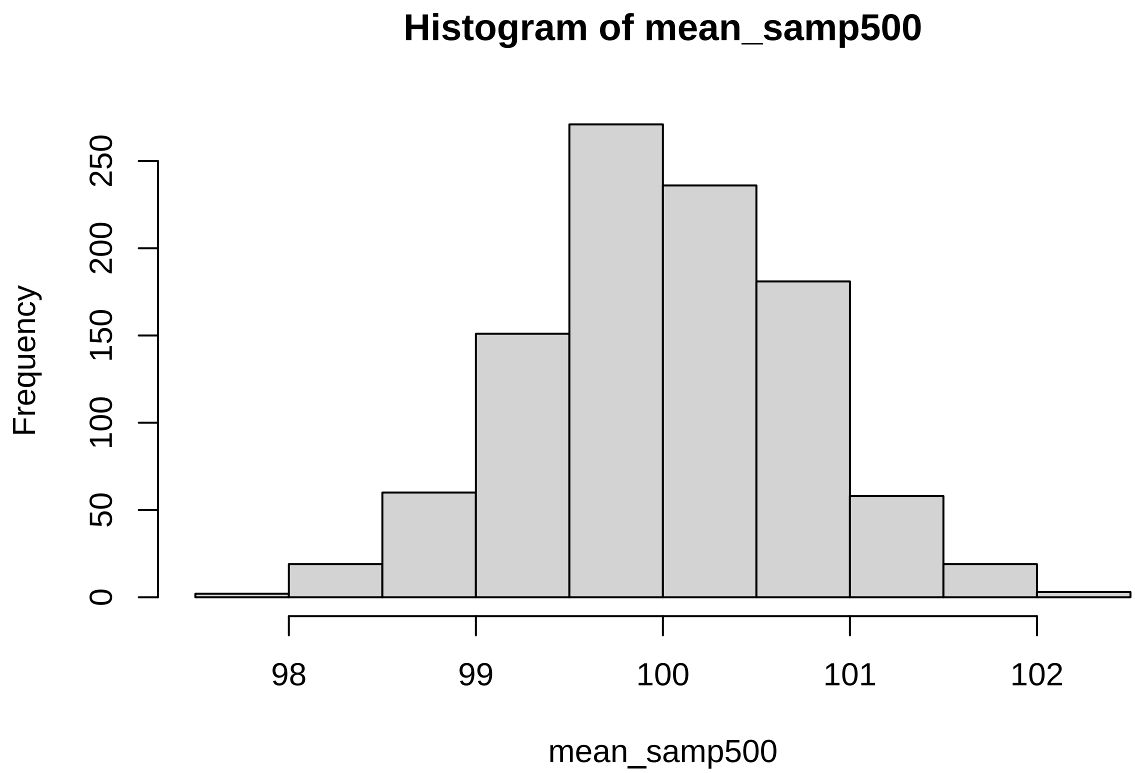

```r
# Sample size = 100
mean_samp100 = numeric(1000)
sd_samp100 = numeric(1000)
for (ii in 1:1000) {
  samp100= rnorm(100,100,15)
  mean_samp100[ii] = mean(samp100)
  sd_samp100[ii] = sqrt(sum((samp100-mean(samp100))^2)/ length(samp100))
}

hist(mean_samp100)
```



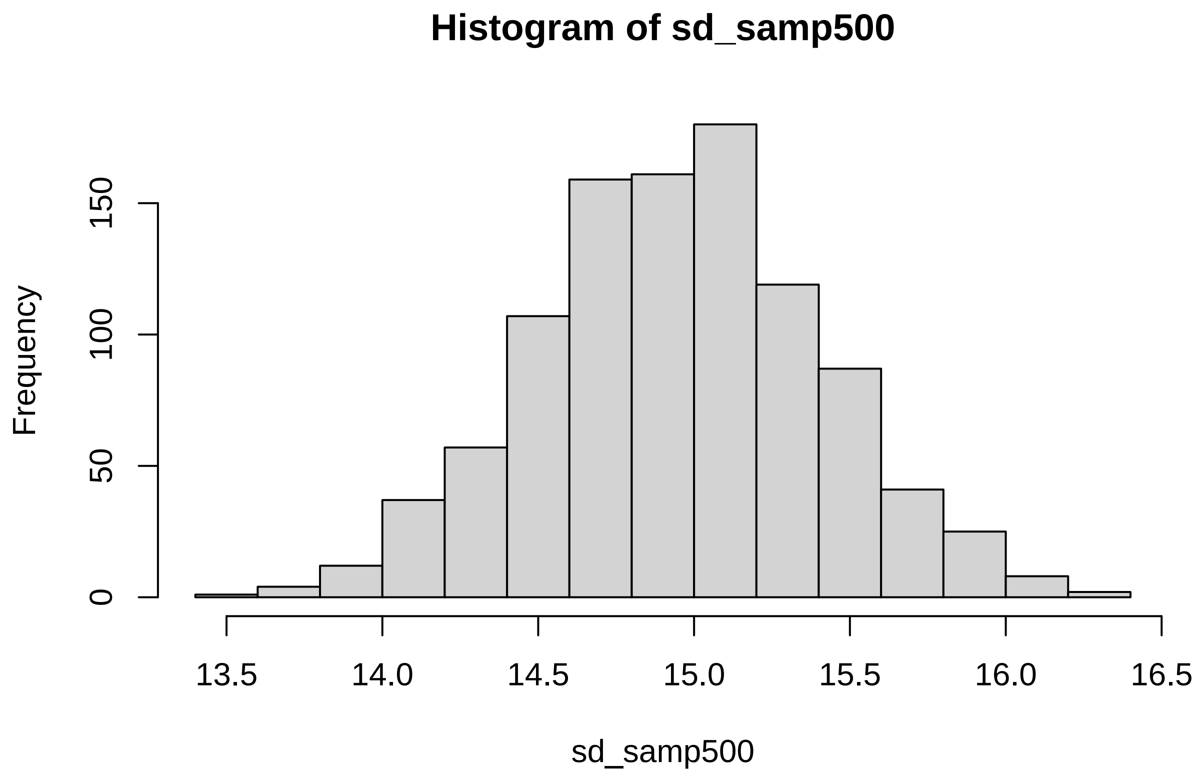

```
hist(sd_samp100)
```



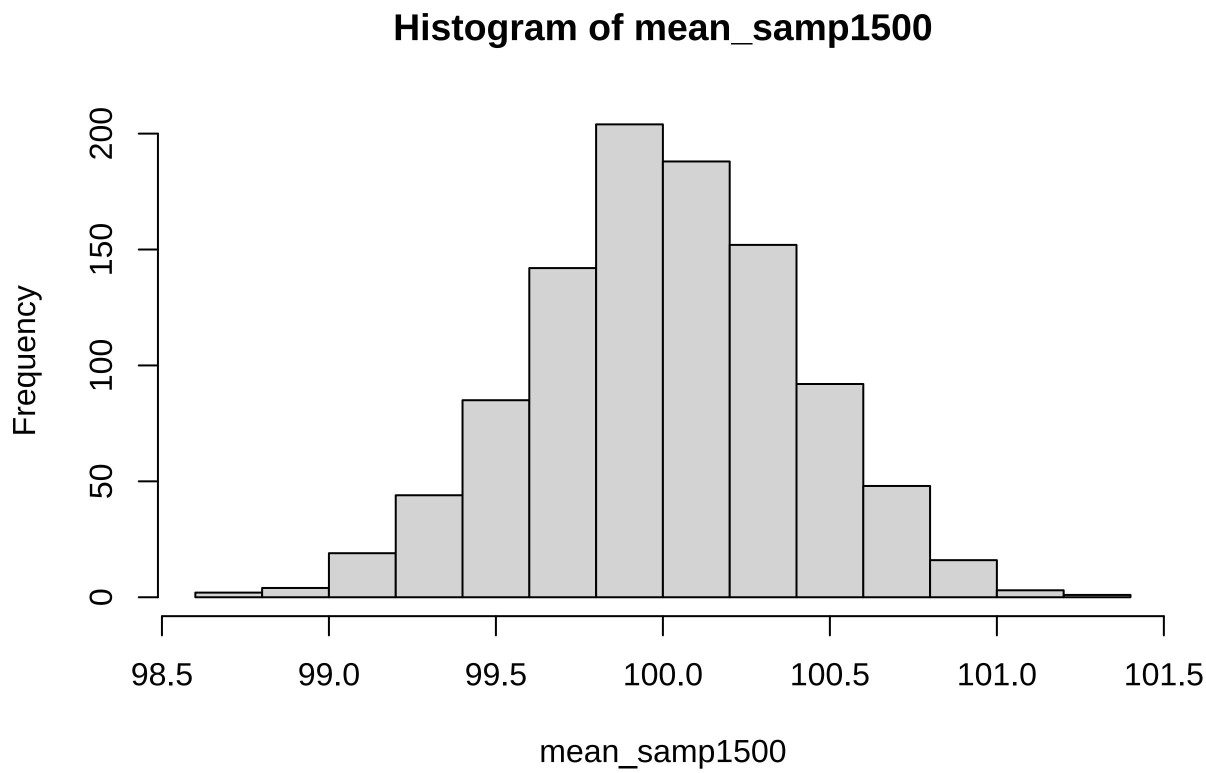

```
# Sample size = 500
mean_samp500 = numeric(1000)
sd_samp500 = numeric(1000)
for (ii in 1:1000) {
  samp500= rnorm(500,100,15)
  mean_samp500[ii] = mean(samp500)
  sd_samp500[ii] = sqrt(sum((samp500-mean(samp500))^2)/ length(samp500))
}

hist(mean_samp500)
```



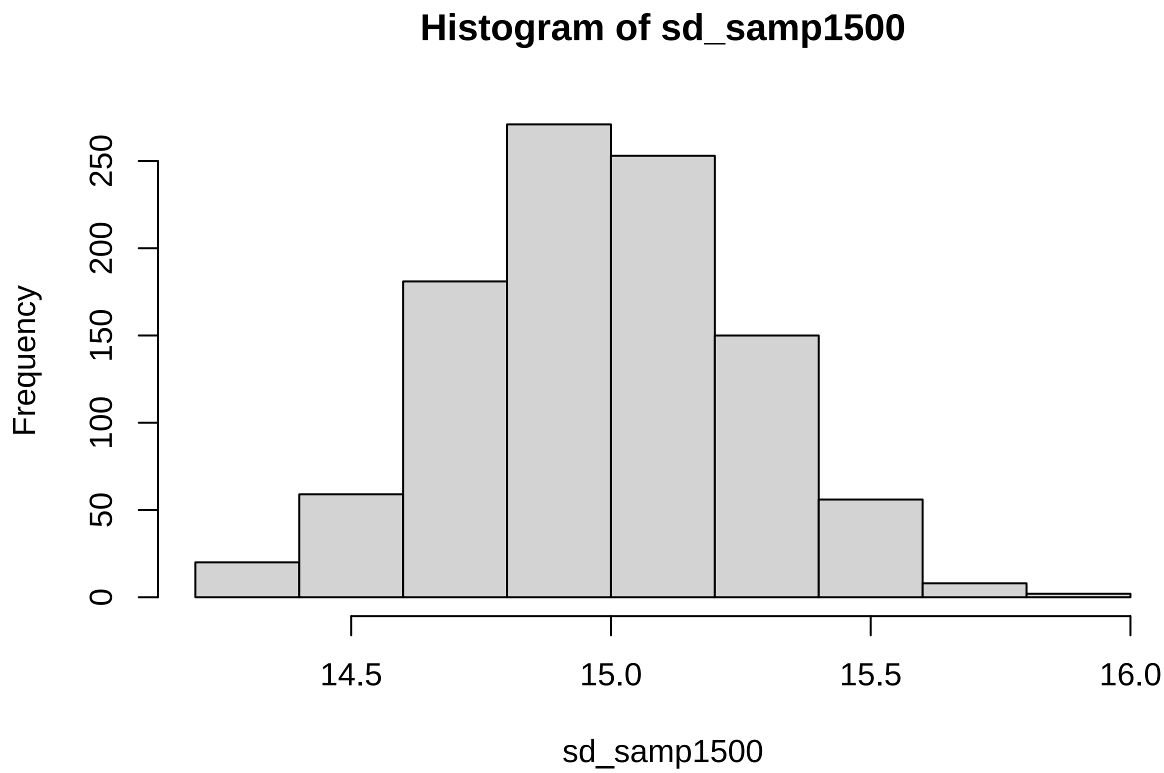

```r
hist(sd_samp500)
```




```r
# Sample size = 1500
mean_samp1500 = numeric(1000)
sd_samp1500 = numeric(1000)
for (ii in 1:1000) {
  samp1500= rnorm(1500,100,15)
  mean_samp1500[ii] = mean(samp1500)
  sd_samp1500[ii] = sqrt(sum((samp1500-mean(samp1500))^2)/ length(samp1500))
}

hist(mean_samp1500)
```

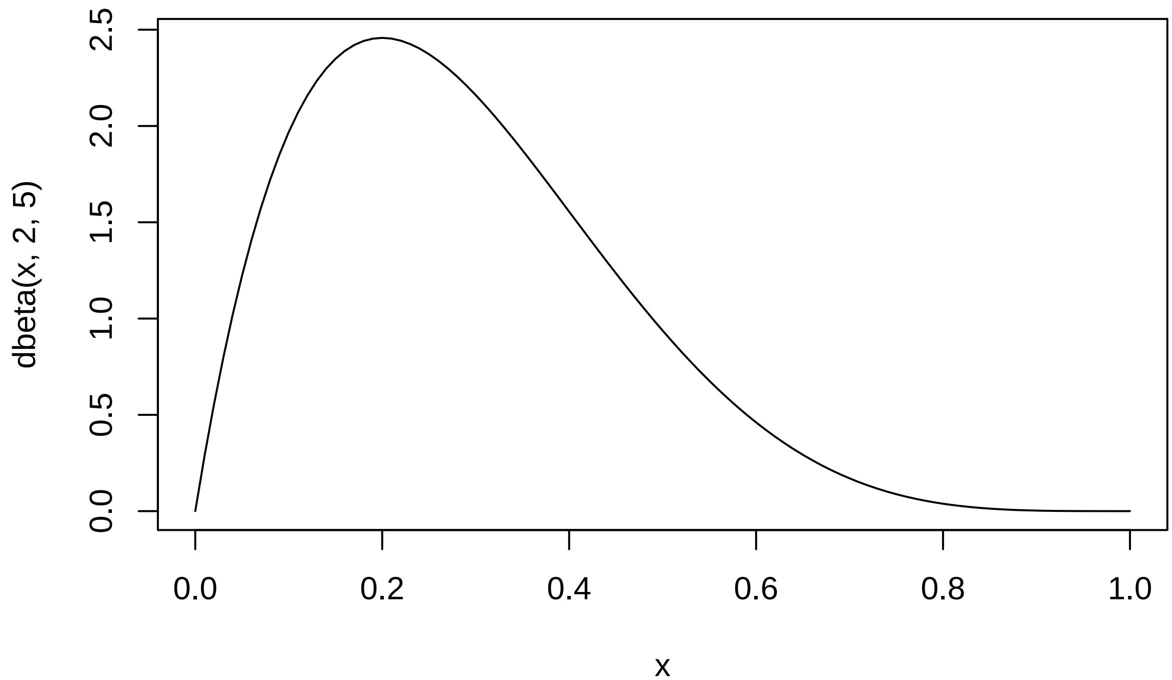



```
hist(sd_samp1500)
```

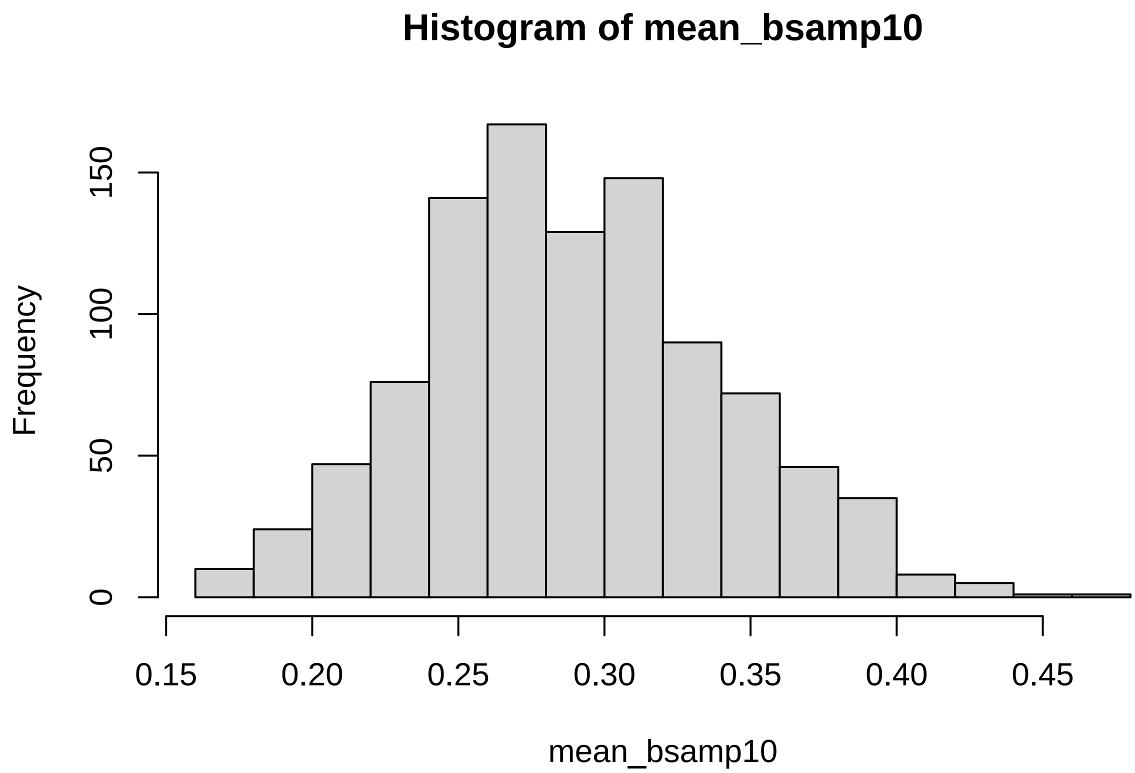


Question 2

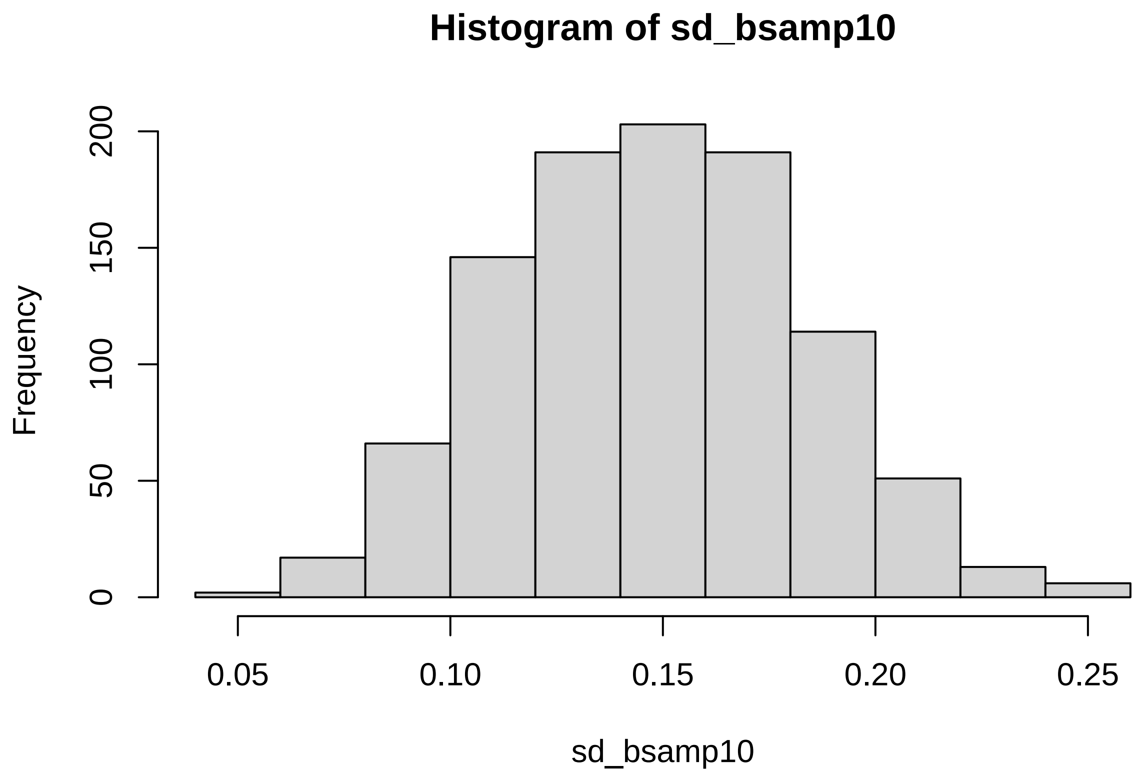
```
x<-seq(0,1,.01)
plot(x, dbeta(x,2,5),type = "l")
```



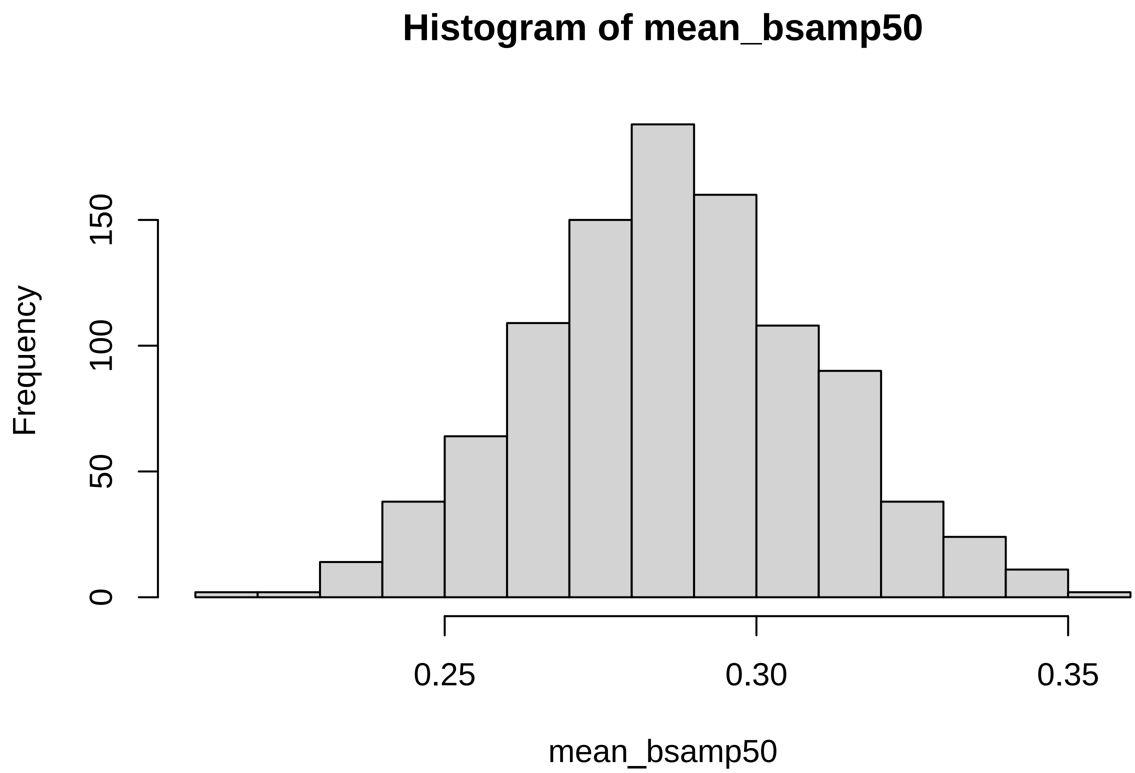

```
mean_bsamp10 = numeric(1000)
sd_bsamp10 = numeric(1000)
for (ii in 1:1000) {
  bsamp10= rbeta(10,2,5)
  mean_bsamp10[ii] = mean(bsamp10)
  sd_bsamp10[ii] = sqrt(sum((bsamp10-mean(bsamp10))^2)/ length(bsamp10))
}

hist(mean_bsamp10)
```



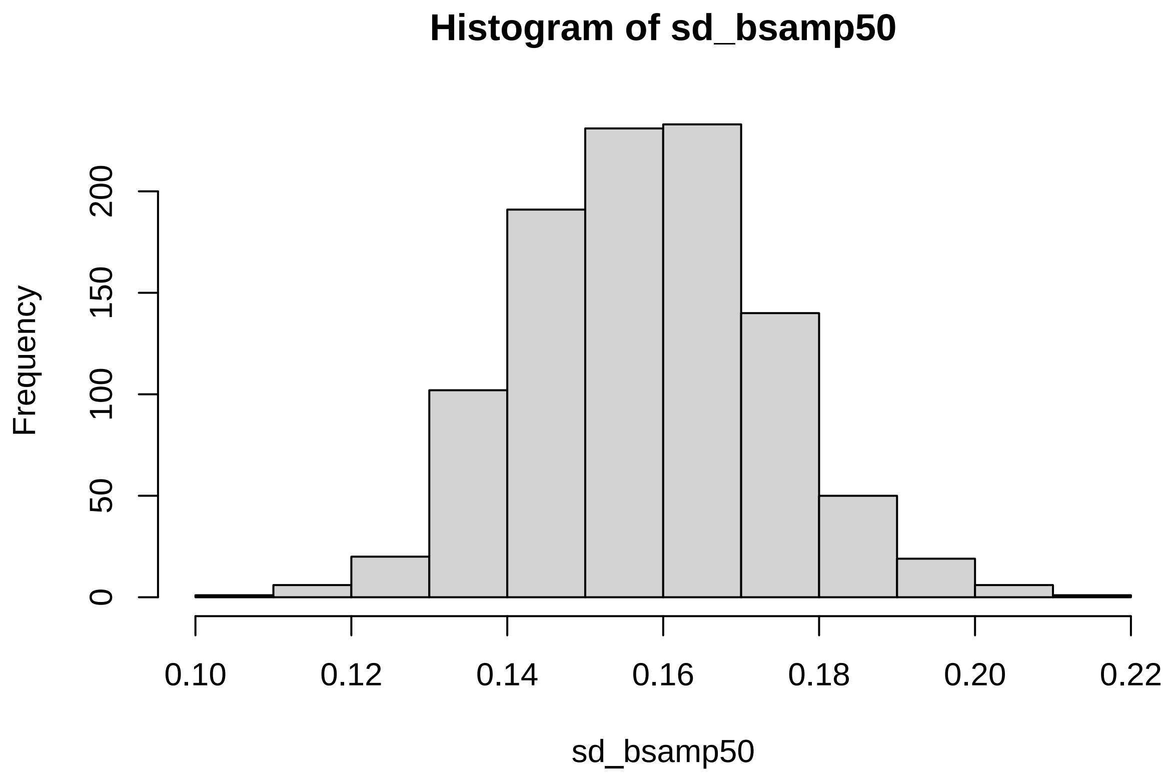

```
hist(sd_bsamp10)
```



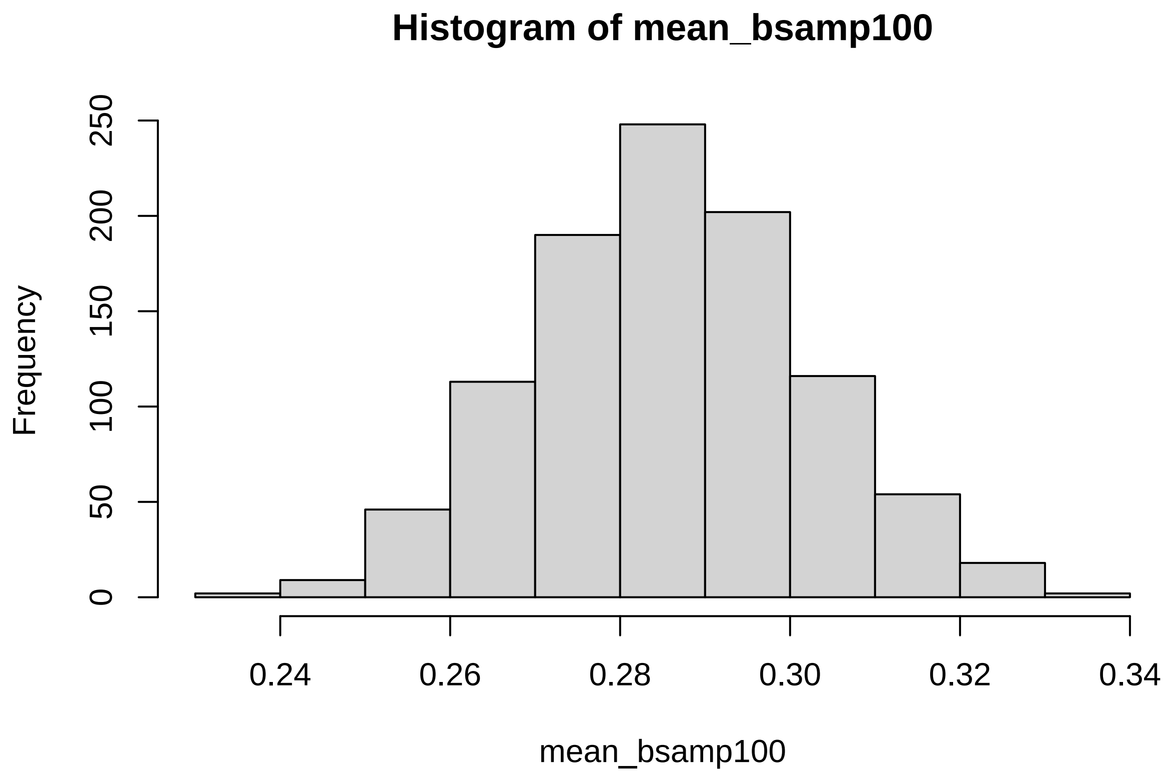

```r
# Sample Size = 50
mean_bsamp50 = numeric(1000)
sd_bsamp50 = numeric(1000)
for (ii in 1:1000) {
  bsamp50= rbeta(50,2,5)
  mean_bsamp50[ii] = mean(bsamp50)
  sd_bsamp50[ii] = sqrt(sum((bsamp50-mean(bsamp50))^2)/ length(bsamp50))
}

hist(mean_bsamp50)
```



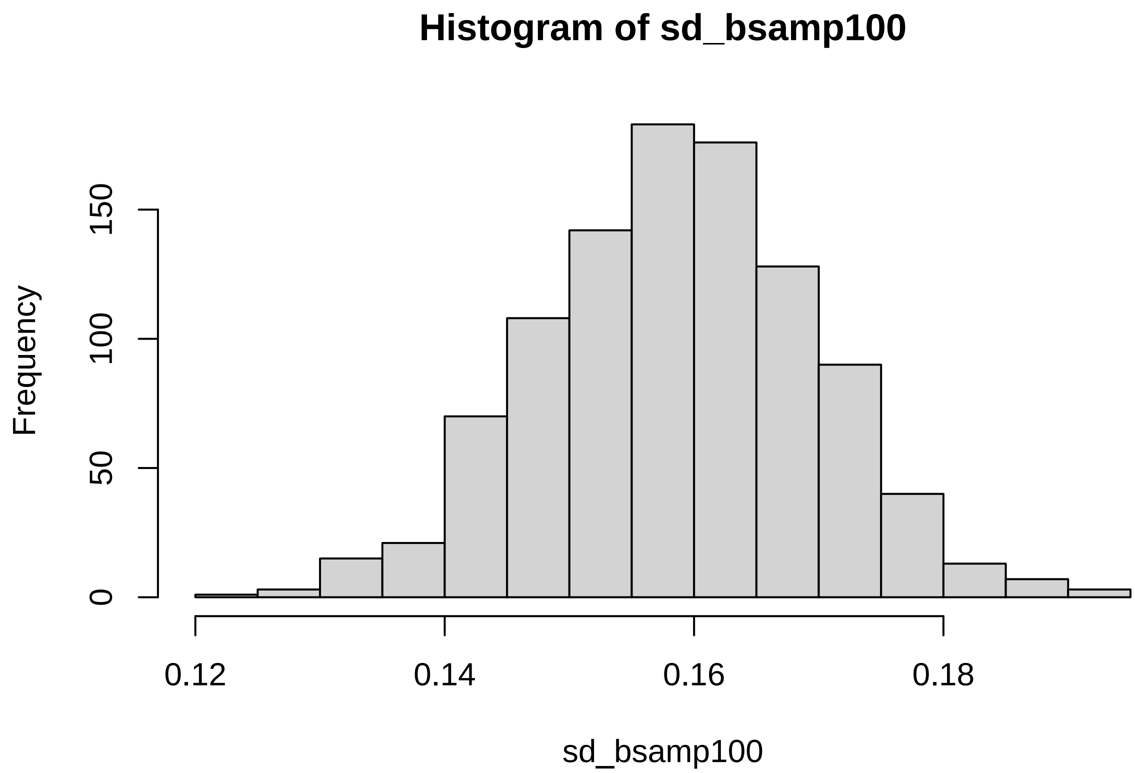

```r
hist(sd_bsamp50)
```



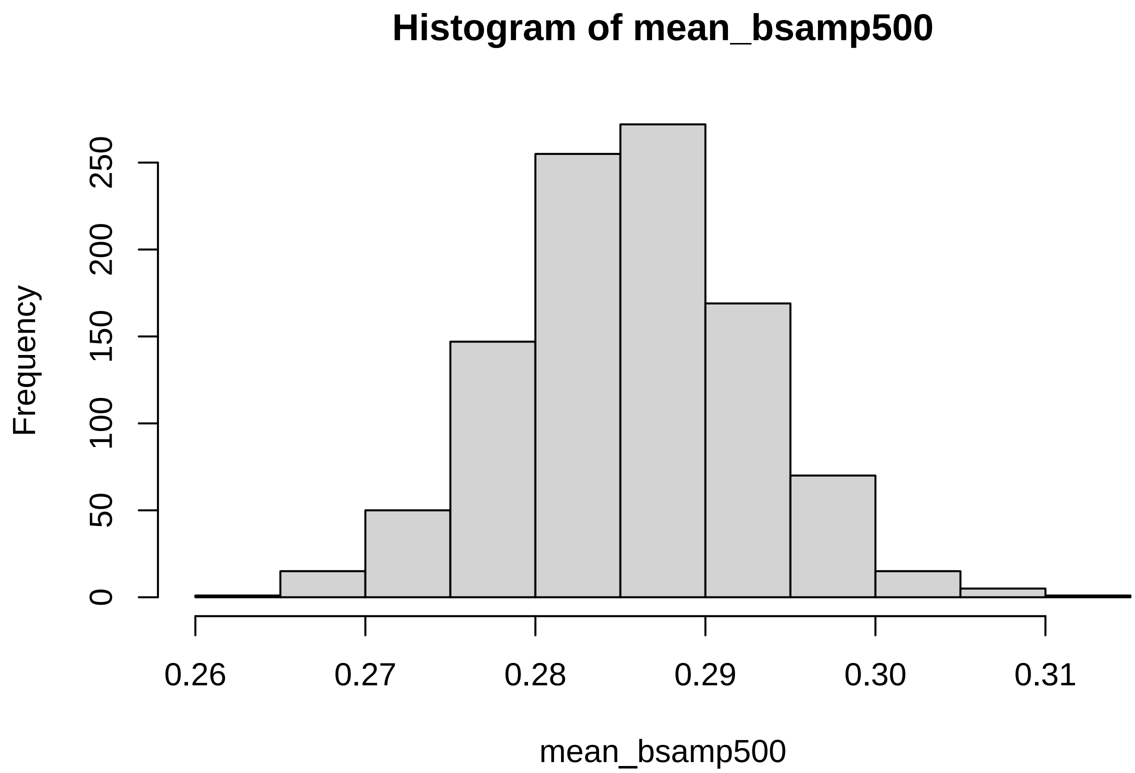

```r
# Sample size = 100
mean_bsamp100 = numeric(1000)
sd_bsamp100 = numeric(1000)
for (ii in 1:1000) {
  bsamp100= rbeta(100,2,5)
  mean_bsamp100[ii] = mean(bsamp100)
  sd_bsamp100[ii] = sqrt(sum((bsamp100-mean(bsamp100))^2)/ length(bsamp100))
}

hist(mean_bsamp100)
```



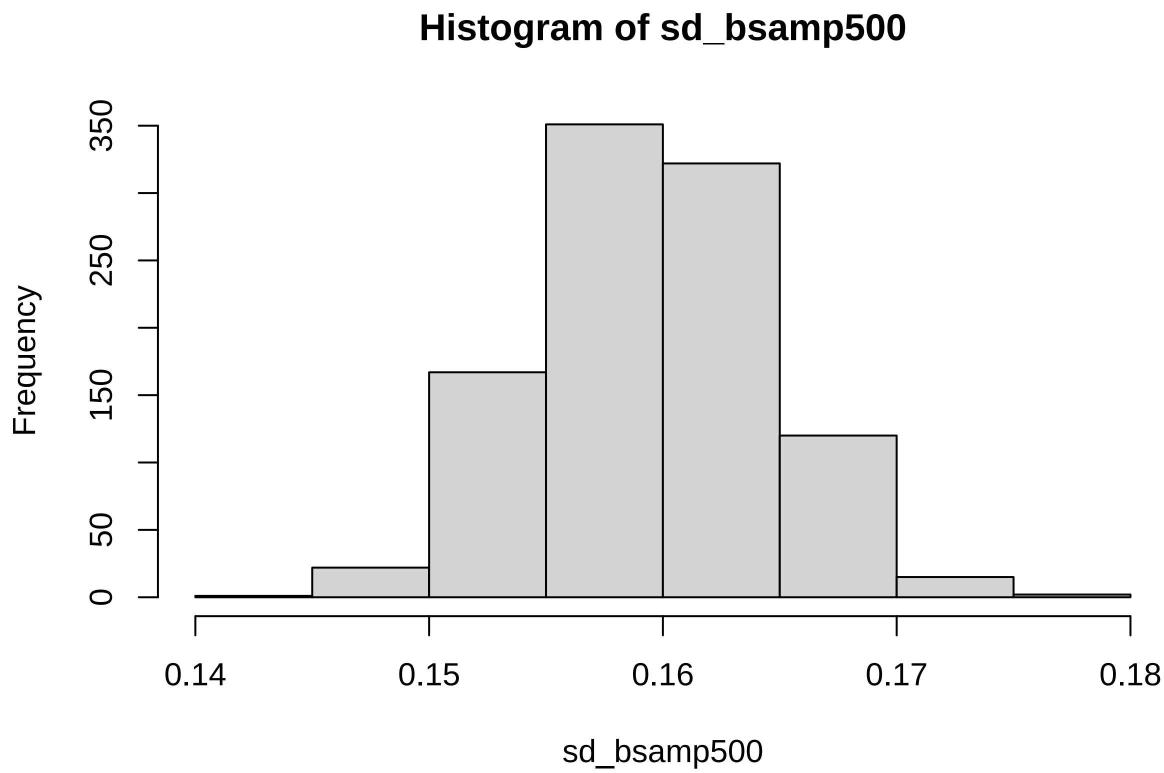

```
hist(sd_bsamp100)
```



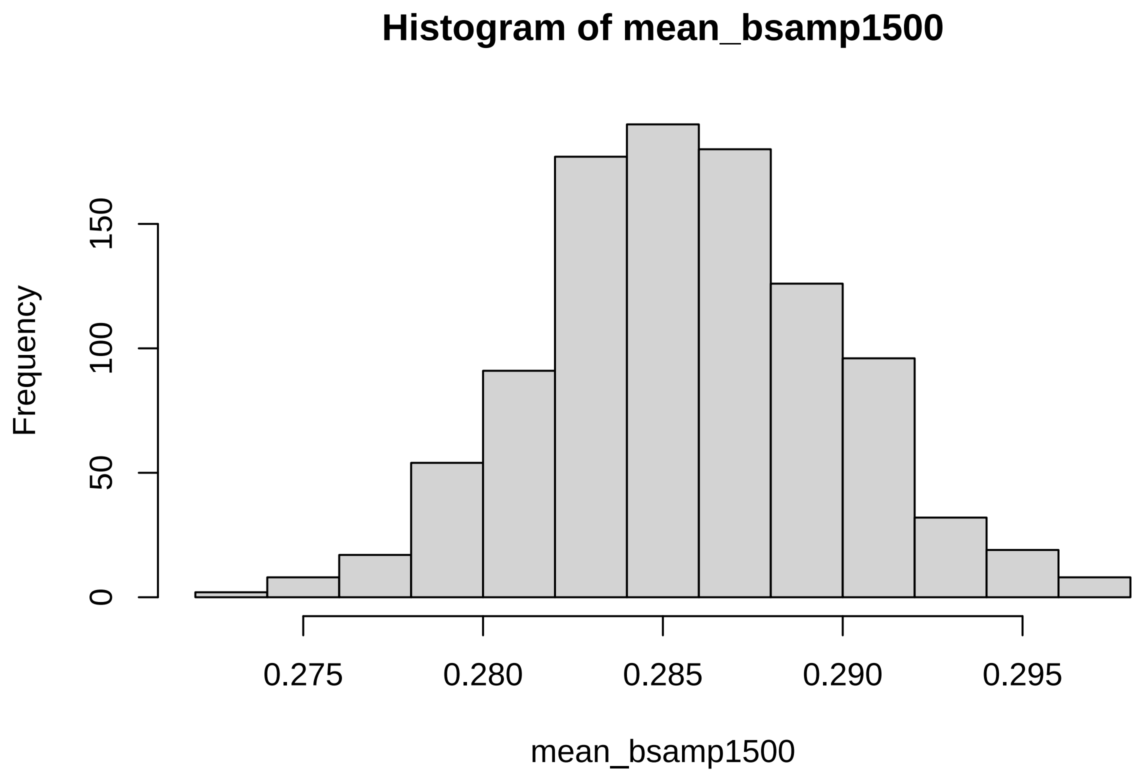

```r
# Sample size = 500
mean_bsamp500 = numeric(1000)
sd_bsamp500 = numeric(1000)
for (ii in 1:1000) {
  bsamp500= rbeta(500,2,5)
  mean_bsamp500[ii] = mean(bsamp500)
  sd_bsamp500[ii] = sqrt(sum((bsamp500-mean(bsamp500))^2)/ length(bsamp500))
}

hist(mean_bsamp500)
```



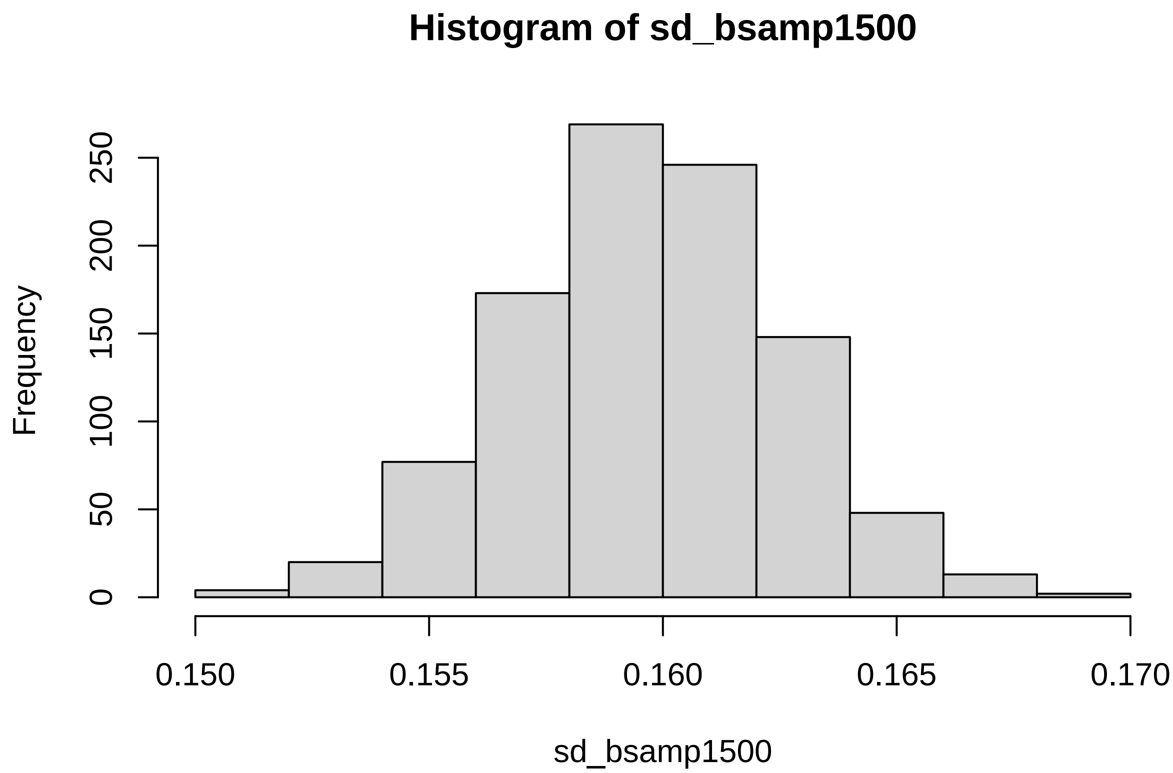

```
hist(sd_bsamp500)
```




```
# Sample size = 1500
mean_bsamp1500 = numeric(1000)
sd_bsamp1500 = numeric(1000)
for (ii in 1:1000) {
  bsamp1500= rbeta(1500,2,5)
  mean_bsamp1500[ii] = mean(bsamp1500)
  sd_bsamp1500[ii] = sqrt(sum((bsamp1500-mean(bsamp1500))^2)/ length(bsamp1500))
}

hist(mean_bsamp1500)
```

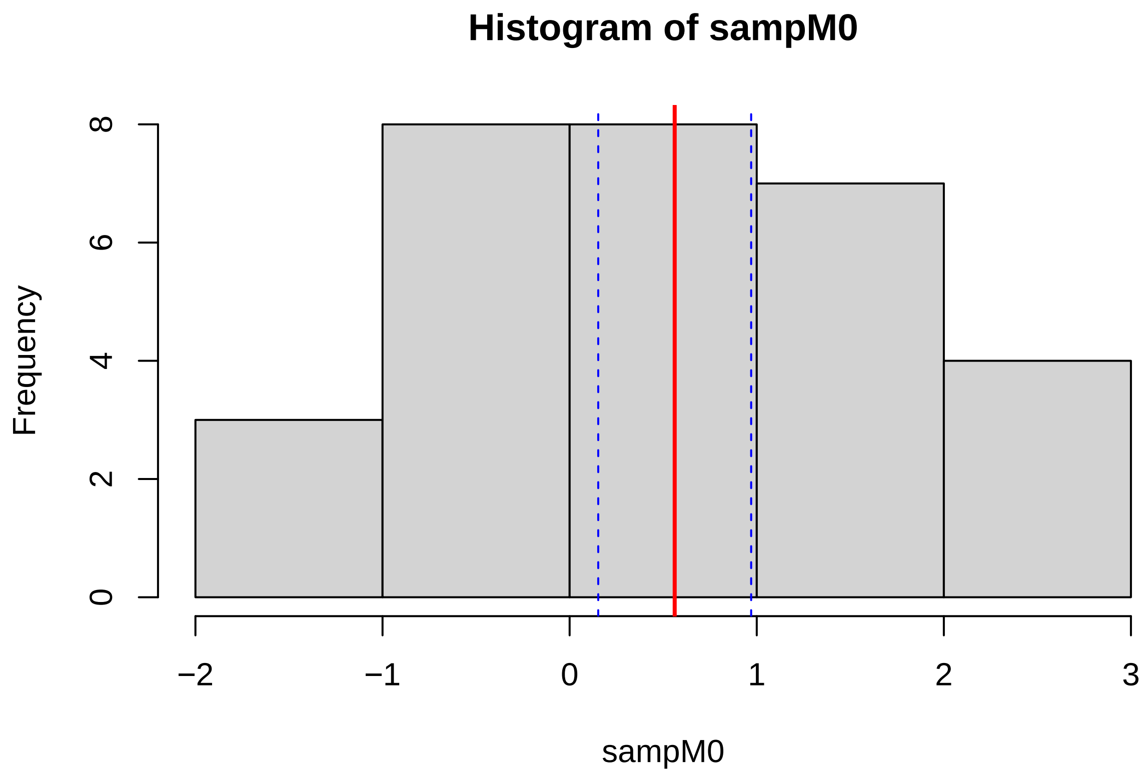



```
hist(sd_bsamp1500)
```

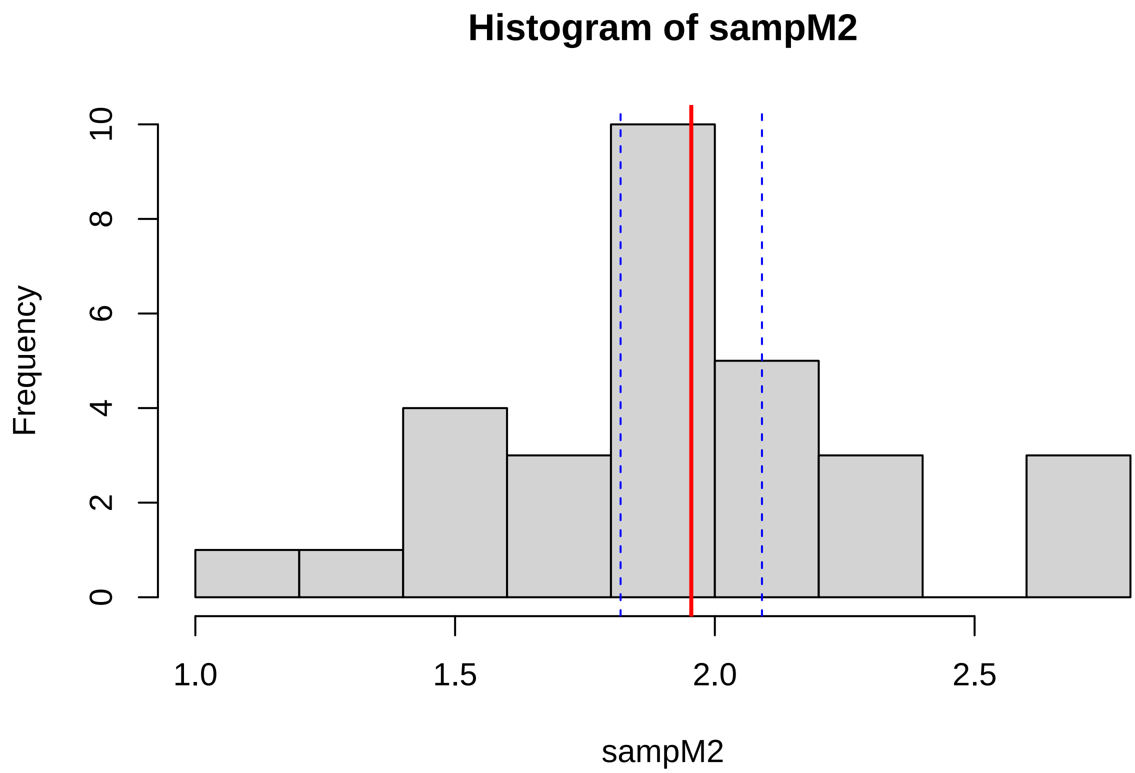


Optional Question 3

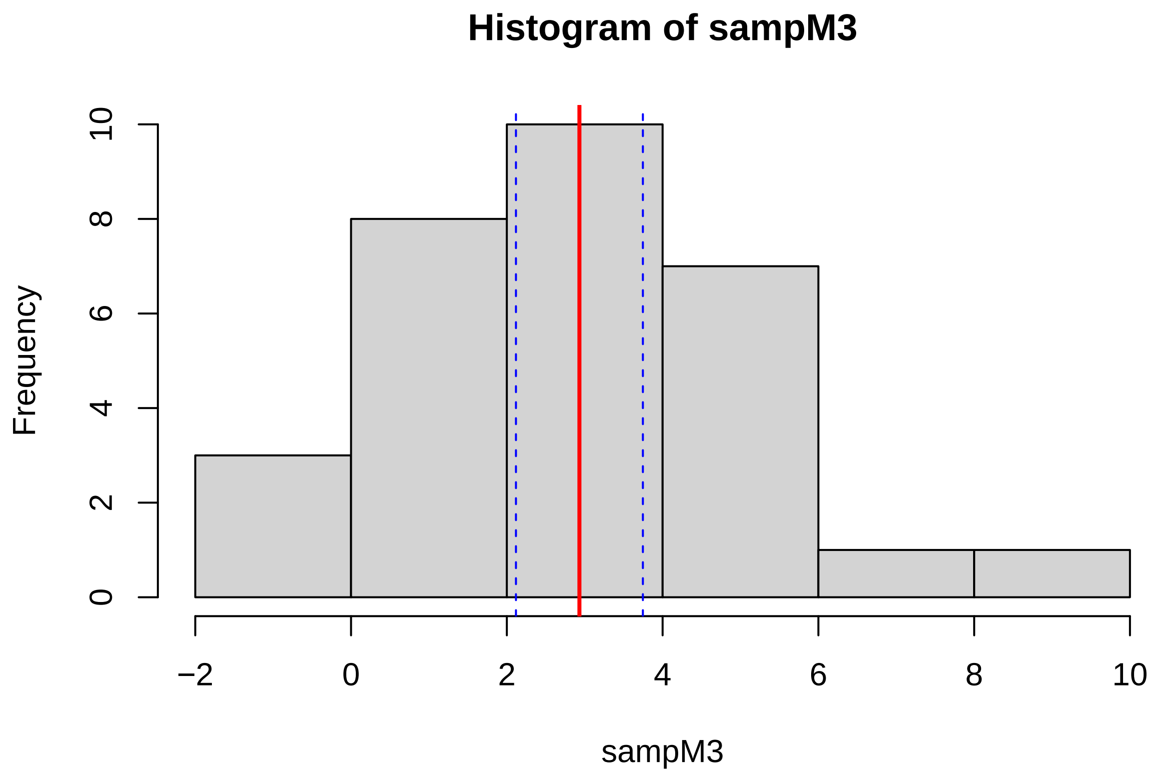
```
sampM0 = rnorm(30,0,1)
sampM2 = rnorm(30,2,0.5)
sampM3 = rnorm(30,3,2)




# Group 1
hist(sampM0)
abline(v = mean(sampM0), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(sampM0) - 1.96*(sd(sampM0)/sqrt(30)), col = "blue", lty = 2)
abline(v = mean(sampM0) + 1.96*(sd(sampM0)/sqrt(30)), col = "blue", lty = 2)
```



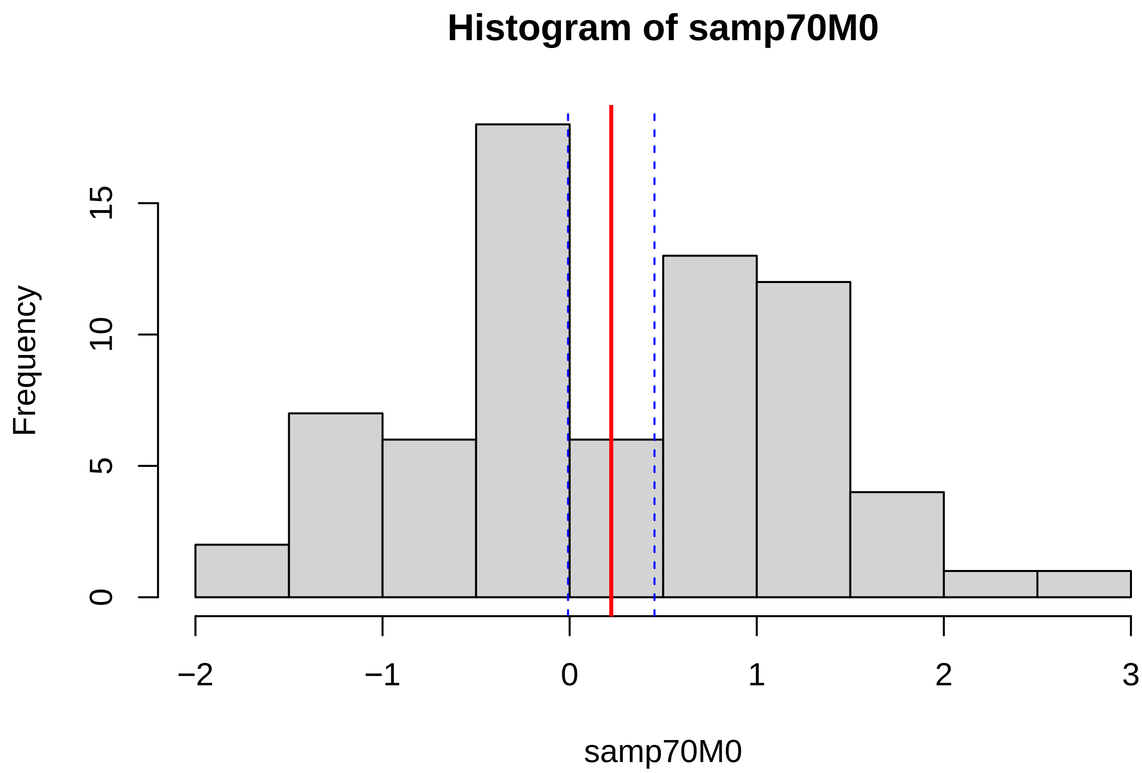

```r
# Group 2
hist(sampM2)
abline(v = mean(sampM2), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(sampM2) - 1.96*(sd(sampM2)/sqrt(30)), col = "blue", lty = 2)
abline(v = mean(sampM2) + 1.96*(sd(sampM2)/sqrt(30)), col = "blue", lty = 2)
```



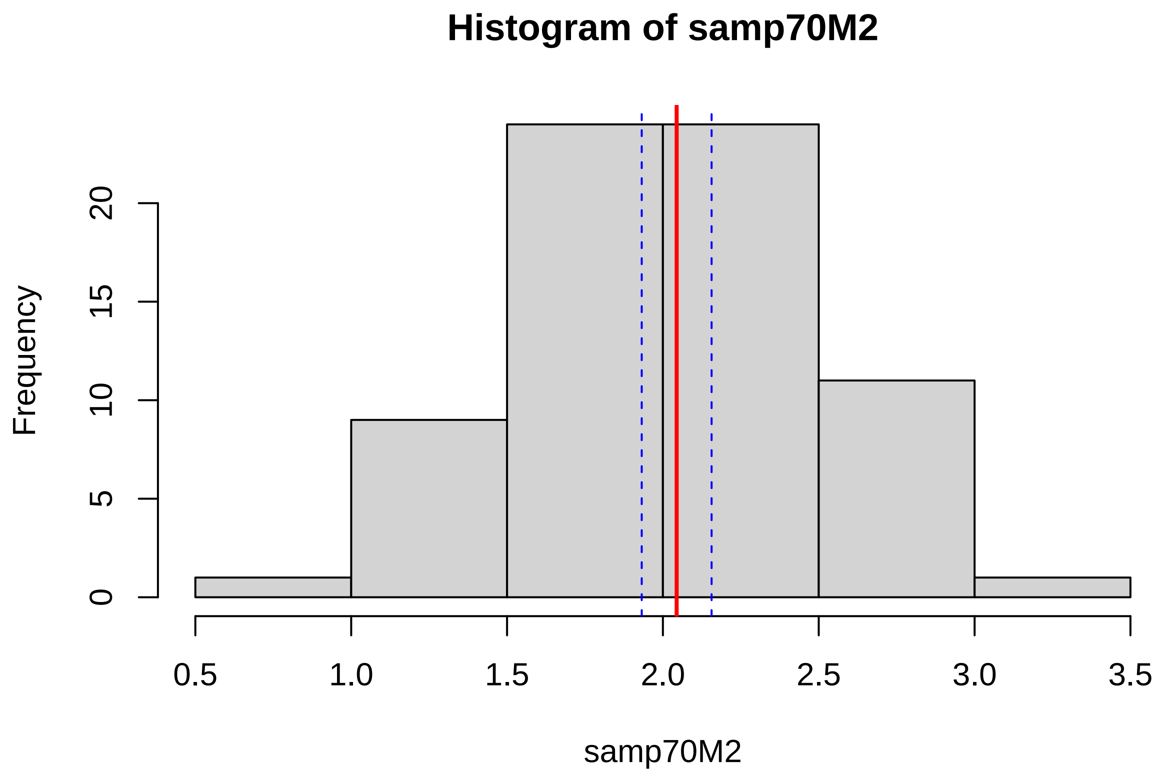

```r
# Group 3
hist(sampM3)
abline(v = mean(sampM3), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(sampM3) - 1.96*(sd(sampM3)/sqrt(30)), col = "blue", lty = 2)
abline(v = mean(sampM3) + 1.96*(sd(sampM3)/sqrt(30)), col = "blue", lty = 2)
```



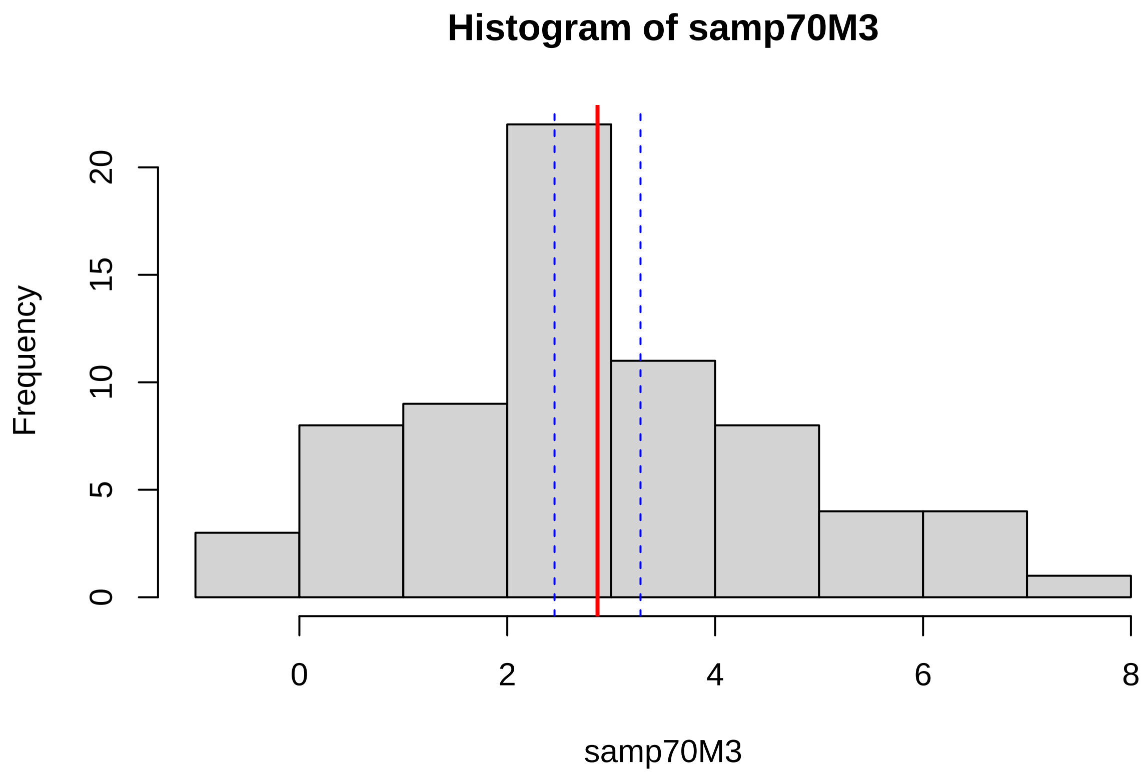

```
# For sample size = 70
samp70M0 = rnorm(70,0,1)
samp70M2 = rnorm(70,2,0.5)
samp70M3 = rnorm(70,3,2)



# Group 1
hist(samp70M0)
abline(v = mean(samp70M0), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp70M0) - 1.96*(sd(samp70M0)/sqrt(70)), col = "blue", lty = 2)
abline(v = mean(samp70M0) + 1.96*(sd(samp70M0)/sqrt(70)), col = "blue", lty = 2)
```



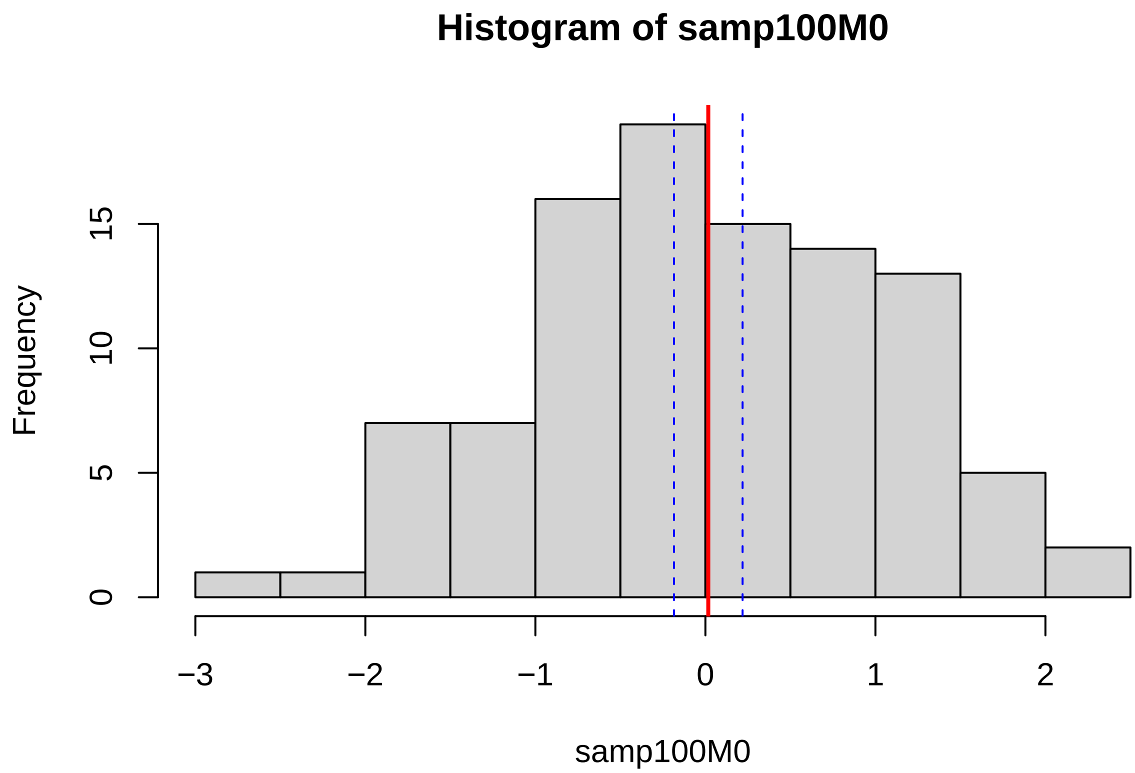

```r
# Group 2
hist(samp70M2)
abline(v = mean(samp70M2), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp70M2) - 1.96*(sd(samp70M2)/sqrt(70)), col = "blue", lty = 2)
abline(v = mean(samp70M2) + 1.96*(sd(samp70M2)/sqrt(70)), col = "blue", lty = 2)
```



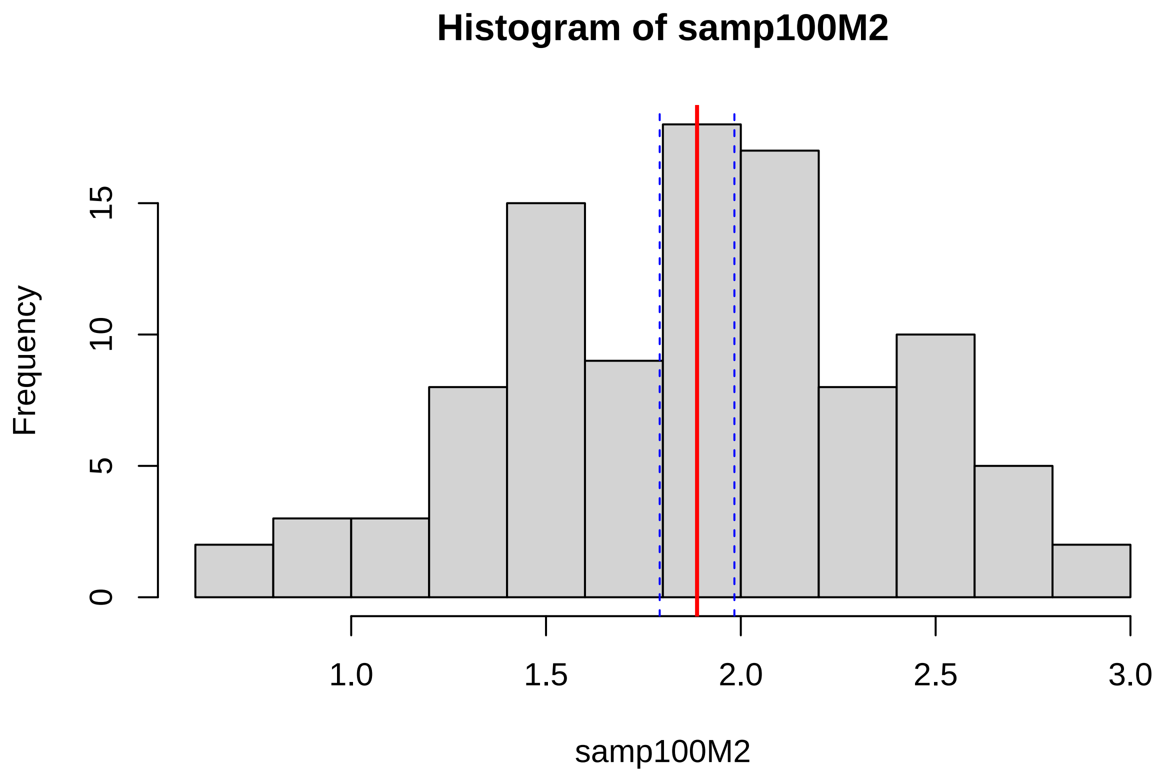

```
# Group 3
hist(samp70M3)
abline(v = mean(samp70M3), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp70M3) - 1.96*(sd(samp70M3)/sqrt(70)), col = "blue", lty = 2)
abline(v = mean(samp70M3) + 1.96*(sd(samp70M3)/sqrt(70)), col = "blue", lty = 2)
```



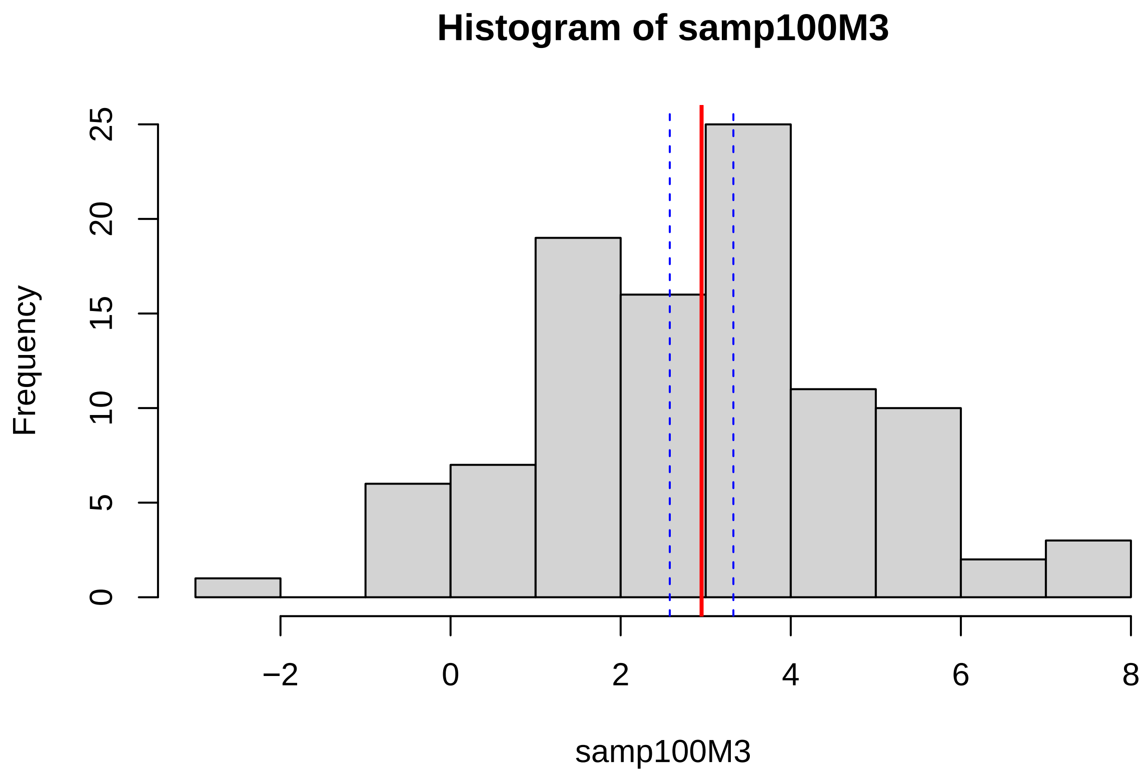

```r
# For sample size = 100
samp100M0 = rnorm(100,0,1)
samp100M2 = rnorm(100,2,0.5)
samp100M3 = rnorm(100,3,2)



# Group 1
hist(samp100M0)
abline(v = mean(samp100M0), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp100M0) - 1.96*(sd(samp100M0)/sqrt(100)), col = "blue",
       lty = 2)
abline(v = mean(samp100M0) + 1.96*(sd(samp100M0)/sqrt(100)), col = "blue",
       lty = 2)
```




```
# Group 2
hist(samp100M2)
abline(v = mean(samp100M2), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp100M2) - 1.96*(sd(samp100M2)/sqrt(100)), col = "blue",
       lty = 2)
abline(v = mean(samp100M2) + 1.96*(sd(samp100M2)/sqrt(100)), col = "blue",
       lty = 2)
```




```
# Group 3
hist(samp100M3)
abline(v = mean(samp100M3), col = "red", lwd = 2) # The Mean
# Confidence Interval lines
abline(v = mean(samp100M3) - 1.96*(sd(samp100M3)/sqrt(100)), col = "blue",
       lty = 2)
abline(v = mean(samp100M3) + 1.96*(sd(samp100M3)/sqrt(100)), col = "blue",
       lty = 2)
```



Interpretations

```
# Question 1 -
# With increasing sample sizes the plots become more like a normal distribution.
# The Standard Deviation of the plots themselves also decreased
# as the sample size increased , i.e. , the Standard error of mean decreased
# with the increase in sample size , depicting the true population mean better
# as guaranteed by the central limit theorem.Same could be seen for the SD plots.

# Question 2 -
# For the Beta distribution, we observe that while the initial population is
# significantly right-skewed, the sampling distribution of the mean becomes
# increasingly symmetric and bell-shaped as the sample size increases. This
# demonstrates the robustness of the Central Limit Theorem, which guarantees
# that the distribution of sample means will converge to normality regardless of
# the population's underlying shape, provided n is sufficiently large. However,
# compared to Question 1, a larger sample size is required here to overcome the
# initial skewness and obtain an accurate estimate of the population mean.


# Question 3 -
# As the number of sampled points increases from 30 to 100, the width of the 95%
# confidence intervals for all three groups consistently narrows. This occurs
# because the margin of error is inversely proportional to the square root of
```

```
# the sample size (n). It can been from the plots that while the intervals for
# the distribution with the largest standard deviation (SD=2) are the widest,
# they still follow the same trend of increasing precision and 'stricter' bounds
# as the sample size increases.
```