PROBABILITY AND STATISTICS

LAB ASSIGNMENT - 8
Based on Central Limit Theorem

Name: Jasween Kaur Brar
Roll No: 102017187
Sub-group: 3CS8

QUES 1) **A pipe manufacturing organization produces different kinds of pipes. We are given the monthly data of the wall thickness of certain types of pipes (data is available on LMS Clt-data.csv).**
The organization has an analysis to perform and one of the basic assumption of that analysis is that the data should be normally distributed.

You have the following tasks to do:
(a) Import the csv data file in R.
(b) Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset.

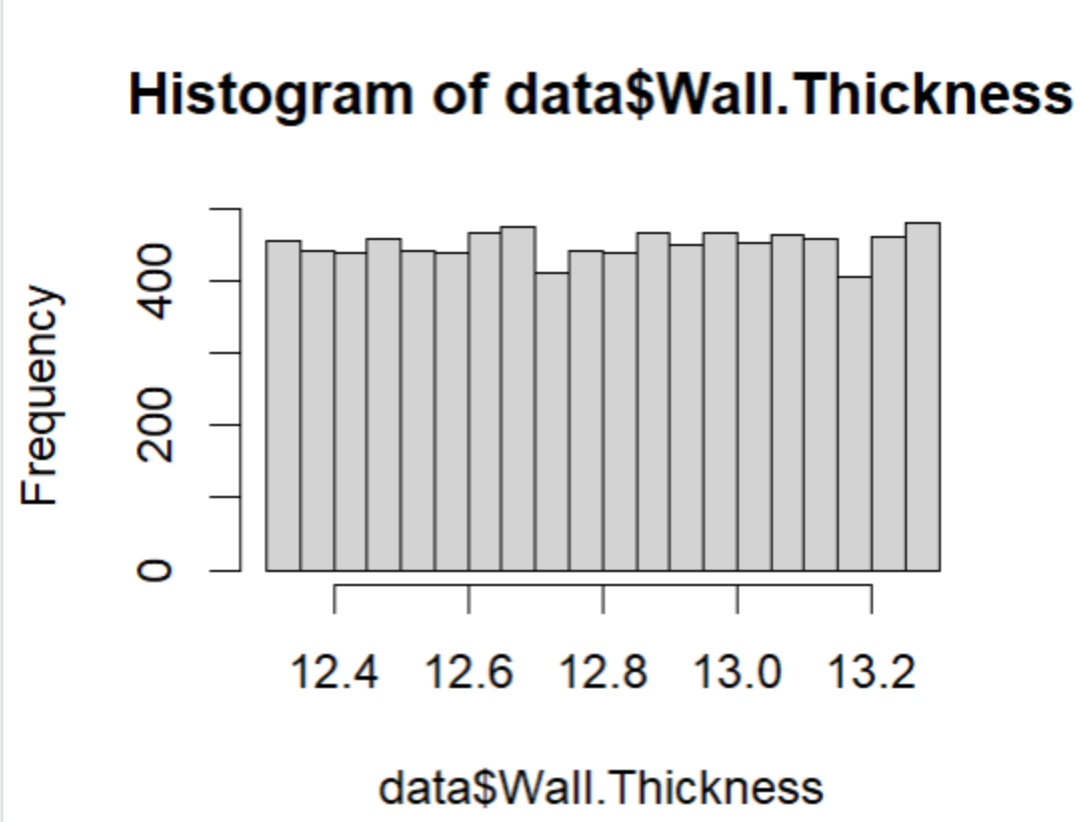
```
# Q1)
# (a) reading of csv file
data <- read.csv(file.choose())
# (b) dimensions
dim(data)
# print top 10 rows
head(data,10)
# or
data[1:10,]
```

```
> dim(data)
[1] 9000    1
> # print top 10 rows
> head(data,10)
   Wall.Thickness
1        12.35487
2        12.61742
3        12.36972
4        13.22335
5        13.15919
6        12.67549
7        12.36131
8        12.44468
9        12.62977
10       12.90381
> # or
> data[1:10,]
 [1] 12.35487 12.61742 12.36972 13.22335 13.15919 12.67549 12.36131 12.44468 12.62977 12.90381
```

(c) Calculate the population mean and plot the observations by making a histogram.

```
# (c)
# mean(data) : will give error
mean(data$Wall.Thickness)
# histogram
hist(data$Wall.Thickness)
```

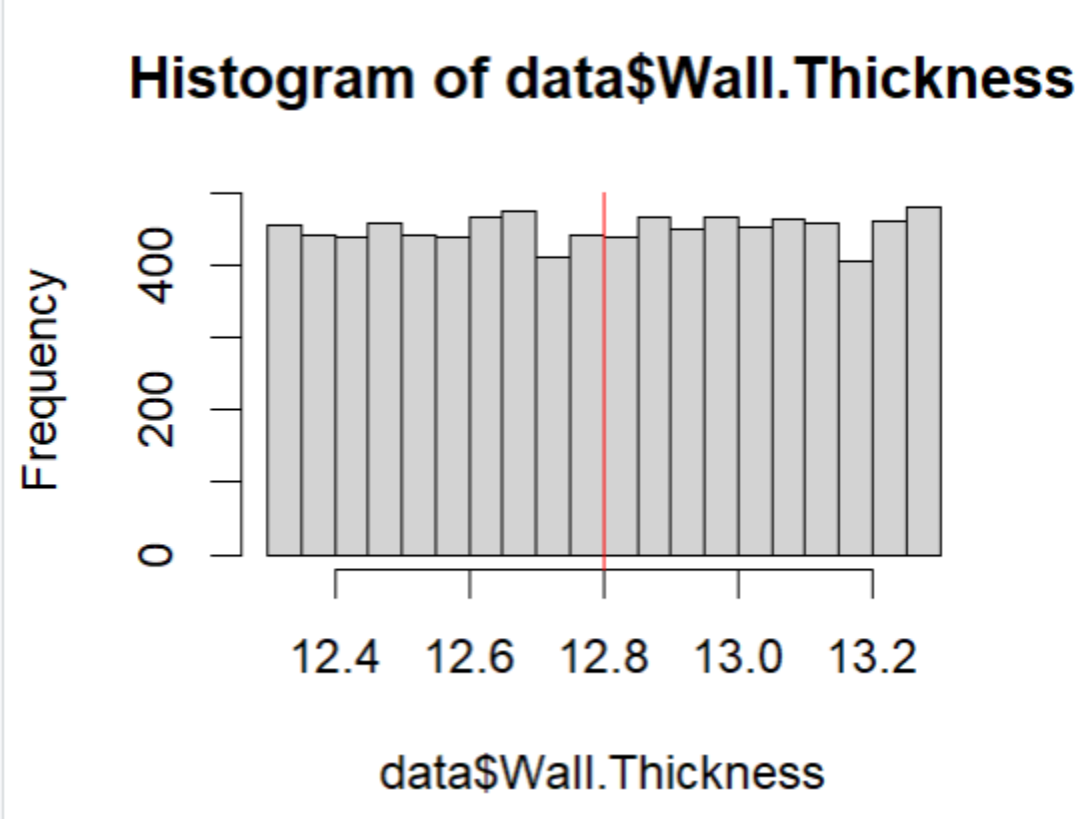
```
> mean(data$Wall.Thickness)
[1] 12.80205
```



(d) Mark the mean computed in last step by using the function abline.

See the red vertical line in the histogram? That's the population mean. Comment on whether the data is normally distributed or not?

```
# (d)
# abline(a,b,h,v,col) -> we need to draw only vertical line to check normal distribution
# Mean  computed using abline :
abline(v=mean(data$Wall.Thickness),col='red')
```



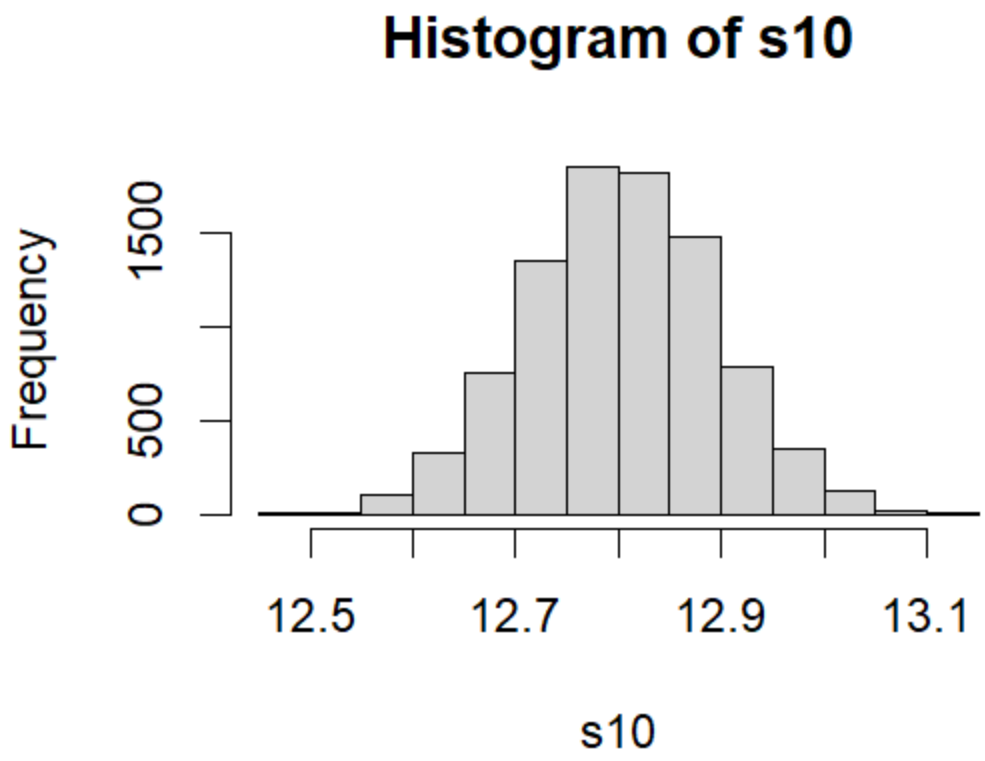
Comment : **The outcome mean divided the data equally this means data is equally distributed. So, data is equally distributed but it does not follow the rules of normal distribution because the histogram doesn't follow bell shaped.**
Max population is not concentrated at the mean value.

Now perform the following tasks:
(a) Draw sufficient samples of size 10, calculate their means, and plot them in R by making histogram. Do you get a normal distribution ?

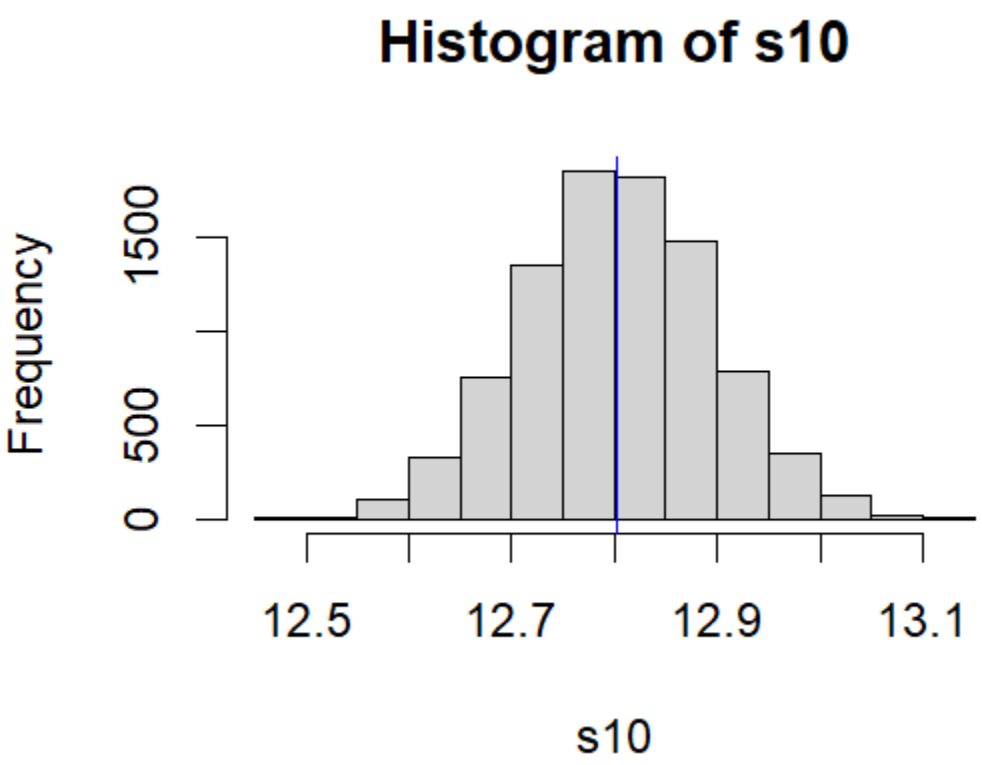
```
# (a)
s10 <- c()
for(i in 1:9000)
{
  s10[i] = mean(sample(data$Wall.Thickness,10,replace=FALSE))
}
m1 <- mean(s10)
m1
```

```
> m1
[1] 12.80272
```

```
hist(s10)
```

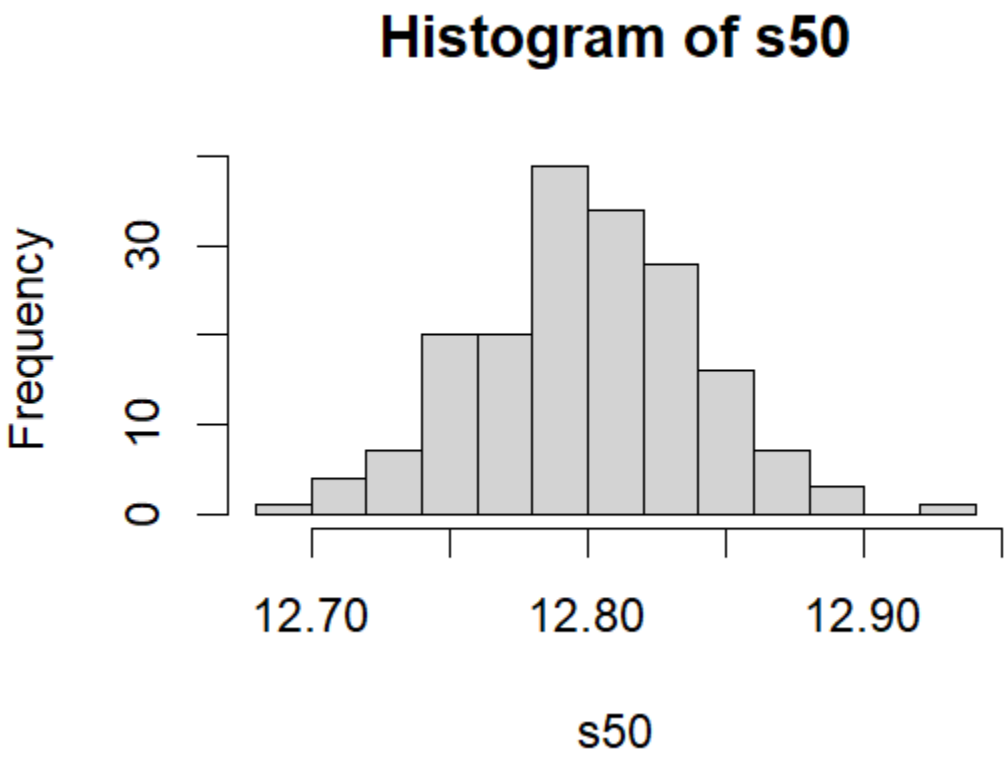



```
abline(v=mean(s10),col='blue')
```

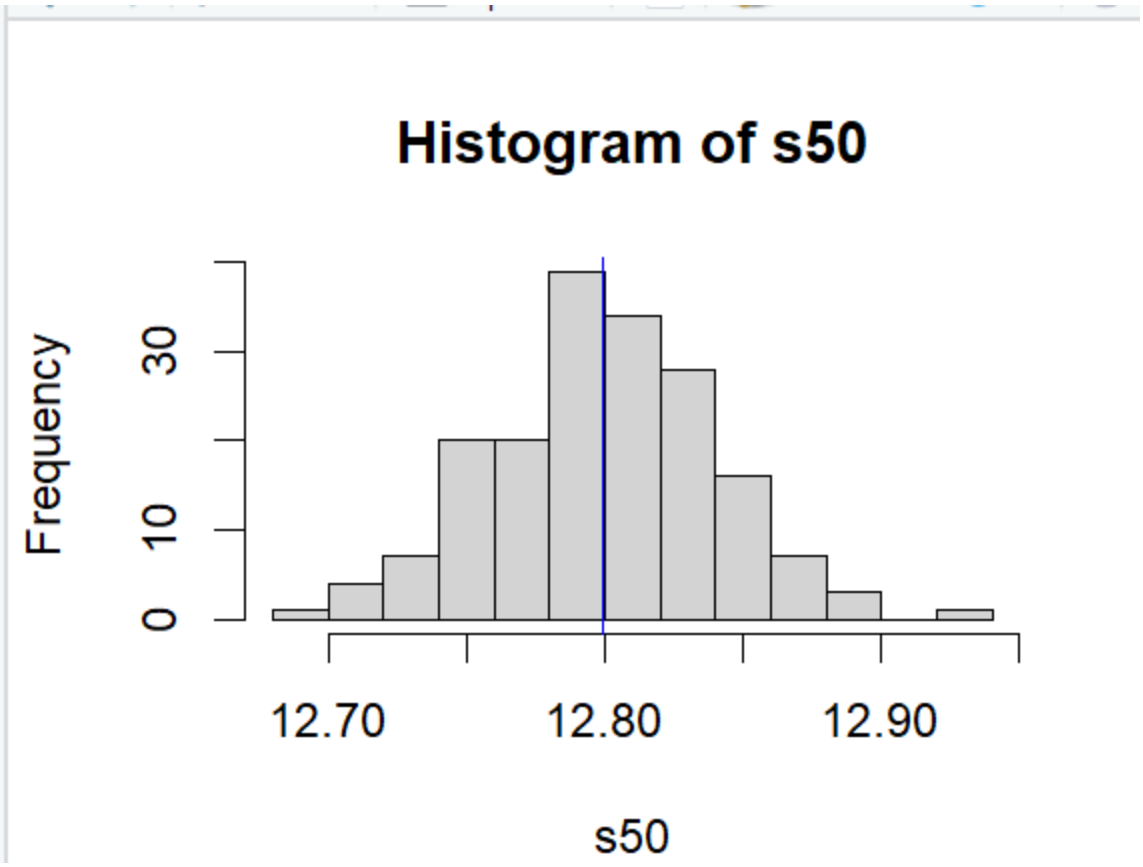


(b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.

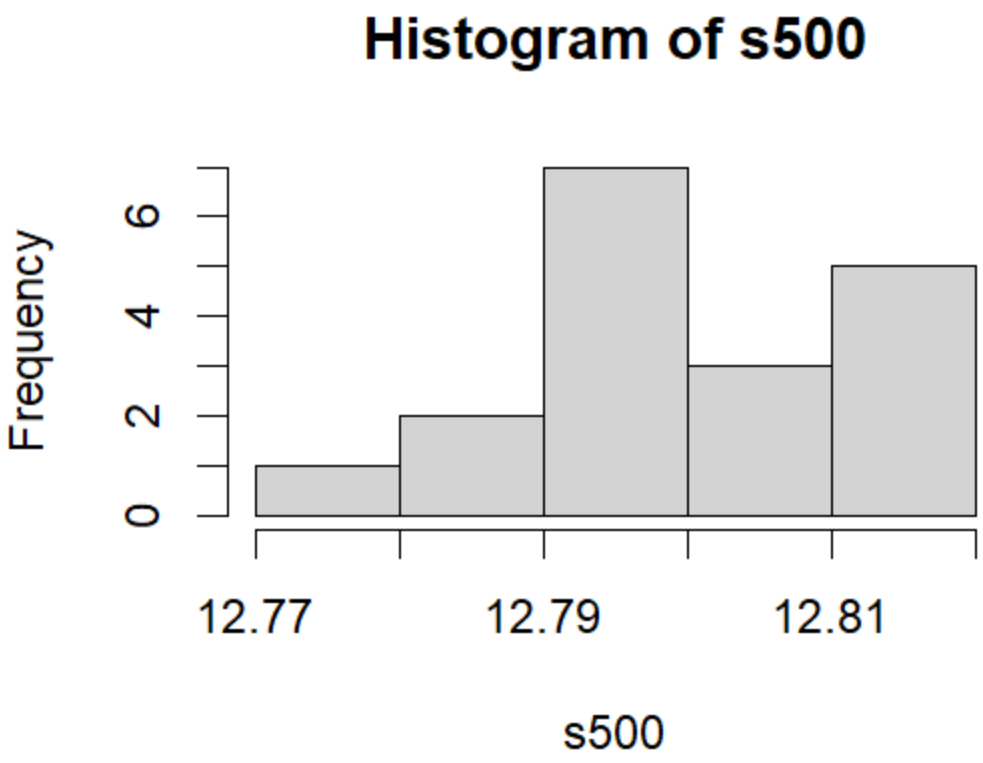
```
# (b)
s50 <- c()
for(i in 1:9000/50)
{
  s50[i] = mean(sample(data$Wall.Thickness,50,replace=FALSE))
}
hist(s50)
```



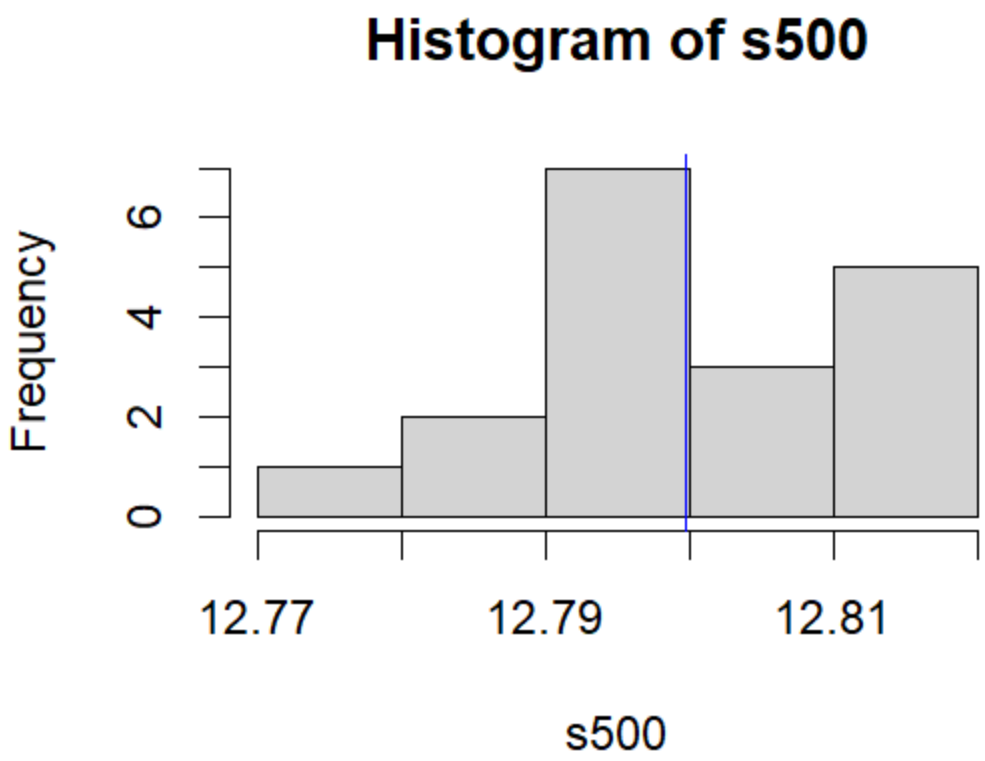

```
abline(v=mean(s50),col='blue')
```



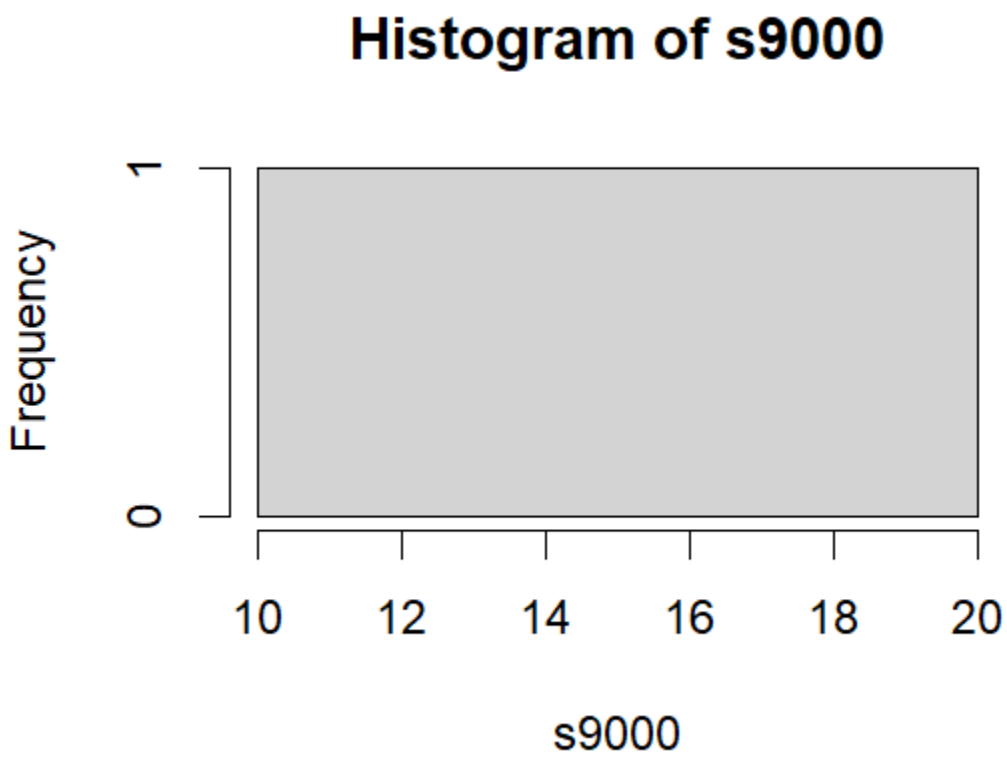

```
# (c)
s500 <- c()
for(i in 1:9000/500)
{
  s500[i] = mean(sample(data$Wall.Thickness,500,replace=FALSE))
}
hist(s500)
```



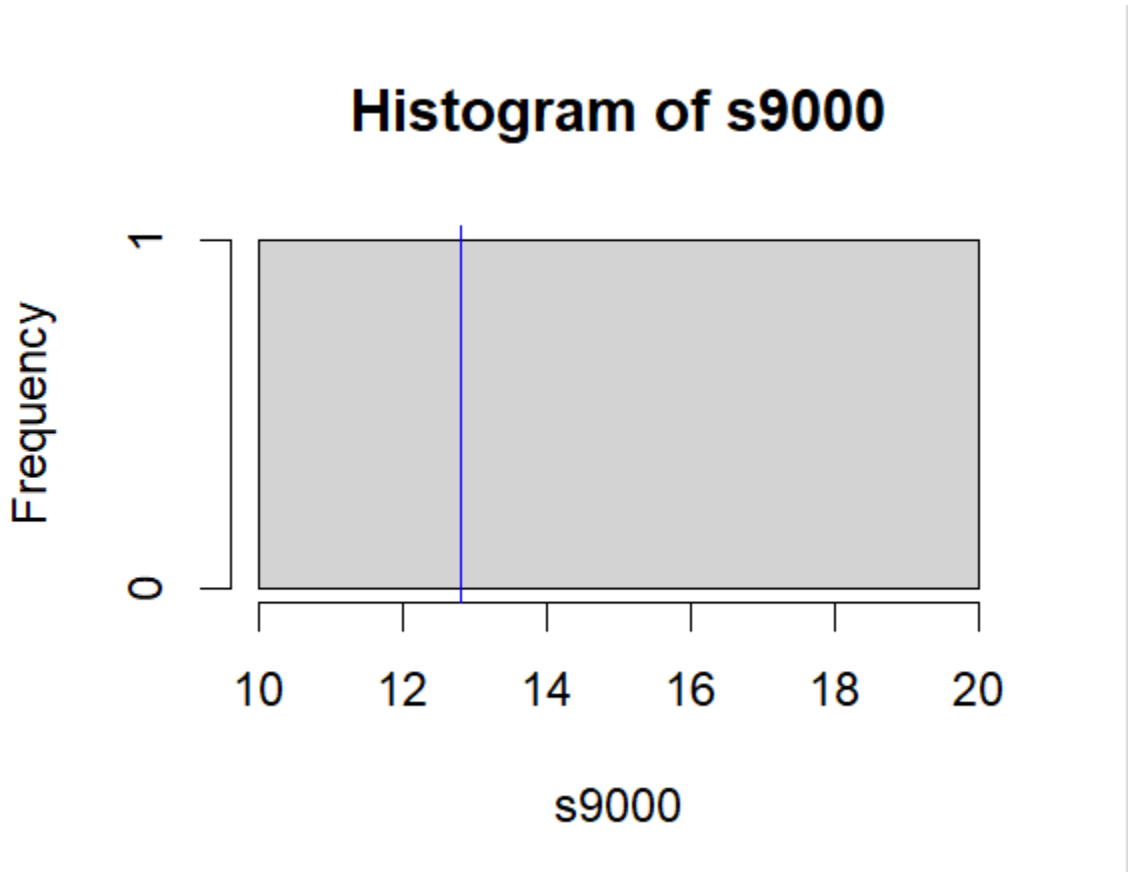

```
abline(v=mean(s500),col='blue')
```




```
# (d)
s9000 <- c()
for(i in 1:9000/9000)
{
  s9000[i] = mean(sample(data$Wall.Thickness,9000,replace=FALSE))
}
hist(s9000)
```




```
abline(v=mean(s9000),col='blue')
```



Here, we get a good bell-shaped curve and the sampling distribution approaches normal distribution as the sample sizes increase. Therefore, we can recommend the organization to use sampling distributions of mean for further analysis.