Lab Assignment-8

Name: Khushi Prasad
SubGroup: 3CO20
Roll No.: 102183044
Subject: Probability & Statistics

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.
CODE:

```
6  data<-read.csv('Clt-data.csv')
```

(b) Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset.
CODE:

```
 9  nrow(data)
10  head(data,10)
```

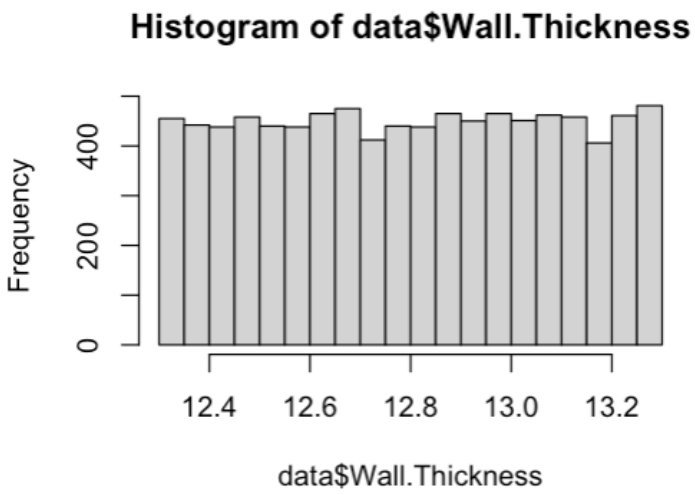
OUTPUT:

```
> nrow(data)
[1] 9000
> 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
```

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

```
13  p<-mean(data$Wall.Thickness)
14  hist(data$Wall.Thickness)
```

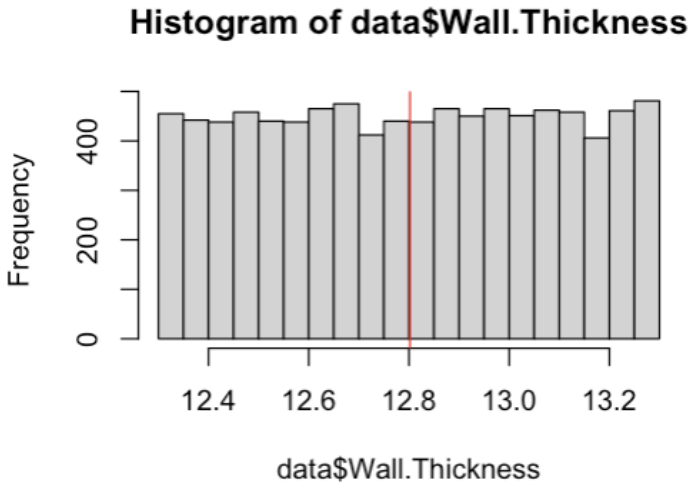
OUTPUT:



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

```
17  abline(v=p, col='red')
```

OUTPUT:

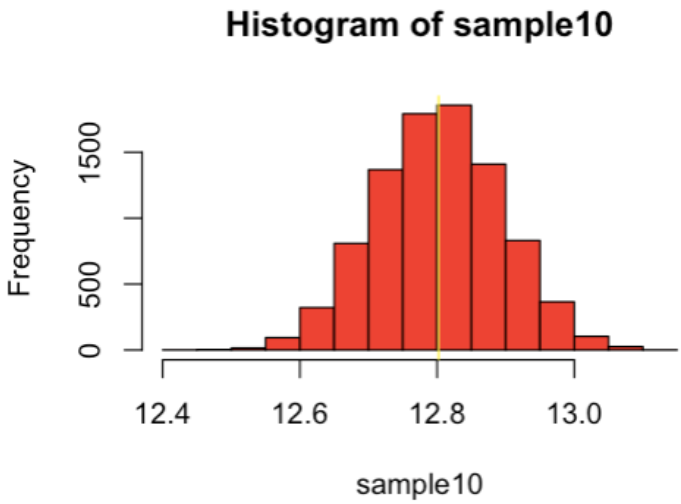


See the red vertical line in the histogram? That's the population mean. Comment on whether the data is normally distributed or not?

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.
CODE:

```
21  sample10<-c()
22  n<-9000
23  for(i in 1:n){
24    sample10[i]<-mean(sample(data$Wall.Thickness,10,replace=TRUE))
25  }
26  hist(sample10,col="red")
27  abline(v=mean(sample10), col="yellow")
```

OUTPUT:



(b) Now repeat the same with sample size 50, 500 and 9000.
CODE:

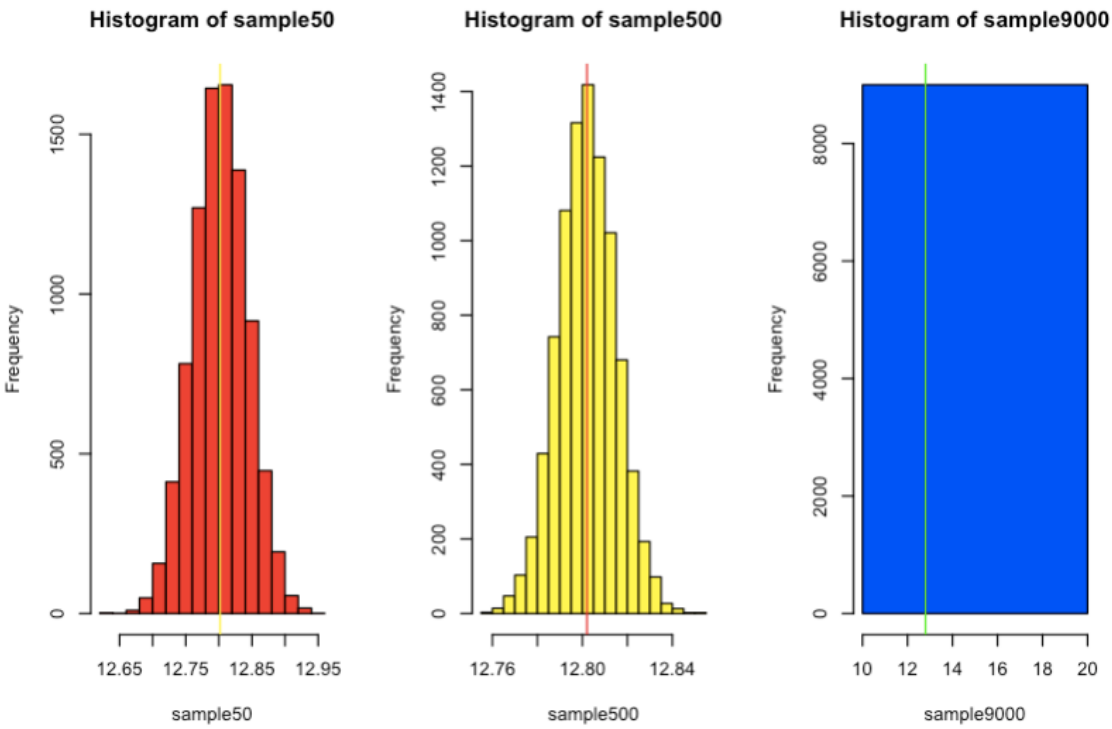
```
sample50<-c()
sample500<-c()
sample9000<-c()
for(i in 1:n){
  sample50[i]<-mean(sample(data$Wall.Thickness,50))
  sample500[i]<-mean(sample(data$Wall.Thickness,500))
  sample9000[i]<-mean(sample(data$Wall.Thickness,9000))
}
par(mfrow=c(1,3))

hist(sample50, col="red")
abline(v=mean(sample50), col="yellow")

hist(sample500,col="yellow")
abline(v=mean(sample500),col="red")

hist(sample9000,col="blue")
abline(v=mean(sample9000),col="green")
```

OUTPUT:



Can you comment on what you observe. 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.