<u>Prob and Stats (Experiment – 8)</u>

102183054
Abha himani

A pipe manufacturing organization produces different kinds of pipes. We are given the monthly data of the wall thickness of certain types of pipes. 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:

data<-read.csv("/home/HP/downloads/Clt-data.csv",header=TRUE, stringAsFactors=FALSE

output:

```
> head(data)
  Wall.Thickness
1       12.35487
2       12.61742
3       12.36972
4       13.22335
5       13.15919
6       12.67549
```

(b)alidate data for correctness by counting number of rows and viewing the top ten rows of the dataset

```
# b) Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset
print(nrow(data))                          # gives no of rows
# ncol(data) : to read the number of columns
head(data, n=10)                           # print top 10 rows
```

output:

```
> print(nrow(data))                        # gives no of rows
[1] 9000
> # ncol(data) : to read the number of columns
> head(data, n=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

```r
# c) calculate the population mean and plot the observations by making a histogram
meann <- mean(data$Wall.Thickness)
print(meann)
hist(data$Wall.Thickness)
```

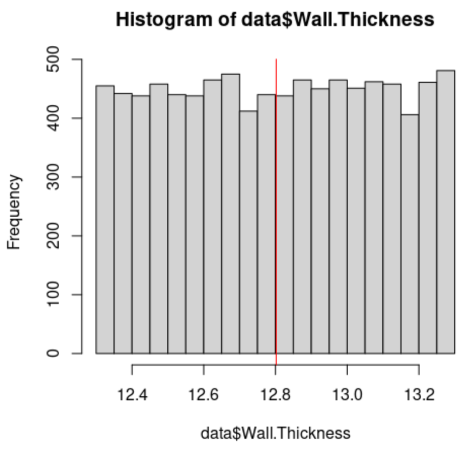
output:

```
> meann <- mean(data$Wall.Thickness)
> print(meann)
[1] 12.80205
> hist(data$Wall.Thickness)
```

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

```r
# d) Mark the mean computed in last step by using the function abline.
abline(v = meann, col = "red")
```

output:



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

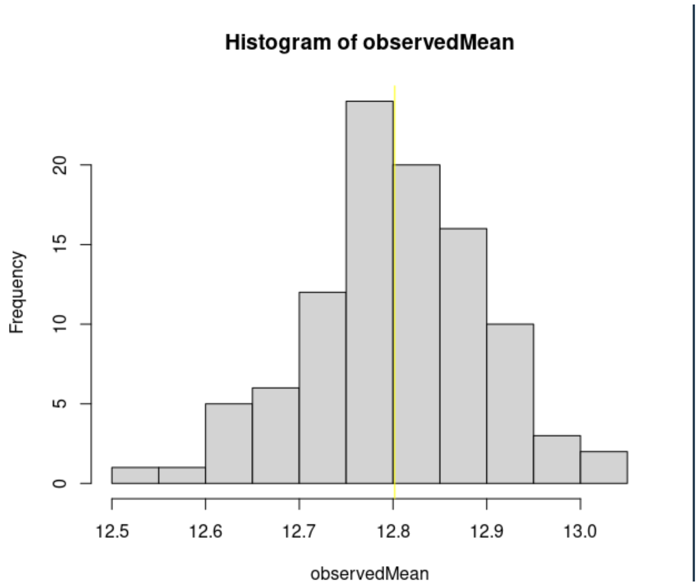
```r
# a) Draw sufficient samples of size 10, calculate their means, and plot them in R by making histogram. Do you get a normal distribution.

observedMean=c()
set.seed(4)
for(x in 1:100){
  data_s1 <- data[sample(1:nrow(data), 10), ]
  observedMean<-append(observedMean,mean(data_s1))

}
print(observedMean)
hist(observedMean)
abline(v=mean(observedMean), col="yellow")
```

output:

```
> print(observedMean)
 [1] 12.84673 12.63678 12.77164 12.78873 12.62353 12.78187 12.74337 12.77175 12.87806 12.82358 12.88800 12.83016 12.66699 12.78645 12.836
[16] 12.79247 12.91426 12.80718 12.91089 12.60569 13.00212 12.85240 12.98114 12.78337 12.90943 12.75418 12.85306 12.76640 12.91014 12.7495
[31] 12.94022 12.88953 12.68788 12.81223 12.73512 12.71975 12.86105 12.84075 12.86257 12.96365 12.83643 12.69878 12.76179 12.87496 12.5839
[46] 12.62071 13.04553 12.88067 12.72499 12.80193 12.75827 12.85547 12.70058 12.50790 12.85467 12.71729 12.81729 12.89495 12.78095 12.7693
[61] 12.75085 12.78515 12.78327 12.77154 12.87076 12.84443 12.88760 12.81736 12.80664 12.88064 12.83857 12.68348 12.81240 12.64543 12.7654
[76] 12.90434 12.73383 12.73614 12.66127 12.94665 12.92856 12.81155 12.93845 12.80518 12.85501 12.70870 12.78268 12.81232 12.97619 12.8049
[91] 12.74200 12.79476 12.67806 12.75553 12.93230 12.75649 12.80117 12.74818 12.78709 12.77454
> hist(observedMean)
```



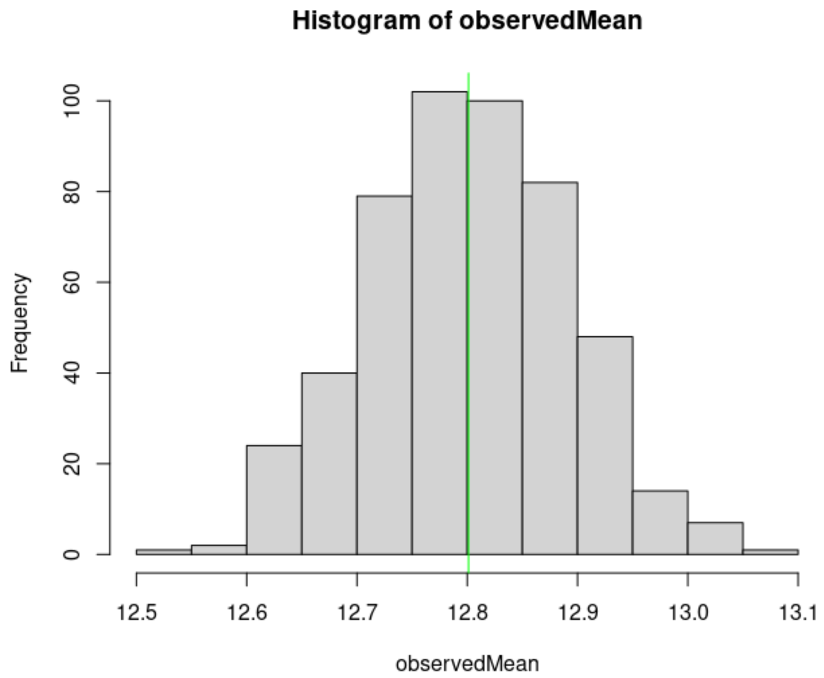
(b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe

```r
# b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.
observedMean=c()
set.seed(4)
for(x in 1:500){
  data_s1 <- data[sample(1:nrow(data), 10), ]
  observedMean<-append(observedMean,mean(data_s1))

}
print(observedMean)
hist(observedMean)
abline(v=mean(observedMean), col="green")
```

output:

```
> print(observedMean)
  [1] 12.84673 12.63678 12.77164 12.78873 12.62353 12.78187 12.74337 12.77175 12.87806 12.82358
 [11] 12.88800 12.83016 12.66699 12.78645 12.83605 12.79247 12.91426 12.80718 12.91089 12.60569
 [21] 13.00212 12.85240 12.98114 12.78337 12.90943 12.75418 12.85306 12.76640 12.91014 12.74957
 [31] 12.94022 12.88953 12.68788 12.81223 12.73512 12.71975 12.86105 12.84075 12.86257 12.96365
 [41] 12.83643 12.69878 12.76179 12.87496 12.58394 12.62071 13.04553 12.88067 12.72499 12.80193
 [51] 12.75827 12.85547 12.70058 12.50790 12.85467 12.71729 12.81729 12.89495 12.78095 12.76934
 [61] 12.75085 12.78515 12.78327 12.77154 12.87076 12.84443 12.88760 12.81736 12.80664 12.88064
 [71] 12.83857 12.68348 12.81240 12.64543 12.76542 12.90434 12.73383 12.73614 12.66127 12.94665
 [81] 12.92856 12.81155 12.93845 12.80518 12.85501 12.70870 12.78268 12.81232 12.97619 12.80493
 [91] 12.74200 12.79476 12.67806 12.75553 12.93230 12.75649 12.80117 12.74818 12.78709 12.77454
[101] 12.64203 13.07983 12.86214 12.89468 12.84903 12.90539 12.90965 12.83944 12.76724 12.90271
[111] 12.73851 12.70032 12.72830 12.73636 12.84785 12.79035 12.85772 12.71896 12.69743 12.70518
[121] 12.94988 12.87664 12.83575 12.75836 12.87031 12.89745 12.65301 12.83435 12.82443 12.63979
[131] 12.72335 12.86066 12.70223 12.79490 12.84064 12.80015 12.85828 12.84422 12.74671 12.87855
[141] 12.81517 12.86865 12.70659 12.71478 12.82118 12.78144 12.65102 12.60667 12.85651 12.80661
[151] 12.77571 12.73260 12.79554 12.83541 12.92273 12.78531 12.71576 12.93238 12.84909 12.80151
[161] 12.71044 12.80648 12.60494 12.81973 12.71776 12.75677 12.66510 12.77014 12.75634 12.84173
[171] 12.64992 12.96773 12.73186 12.91236 12.75614 12.72300 12.79095 12.87012 12.85475 12.70100
[181] 12.75703 12.97464 12.81665 12.97922 12.73572 12.91261 12.87344 12.89955 12.77212 12.96815
[191] 12.83949 12.75080 12.84963 12.78467 12.68395 12.76256 12.73514 12.75944 12.93363 12.65498
[201] 12.79764 12.77493 12.63029 12.75639 12.88146 12.66586 12.76638 12.84454 12.82256 12.80342
[211] 12.73763 12.71977 12.79198 12.69834 12.75846 12.87673 12.89445 12.81395 12.84177 12.68089
[221] 12.85112 12.71303 12.76474 12.76891 12.78963 12.91412 12.73013 12.83470 12.92008 12.81070
[231] 12.95299 12.67884 12.96464 12.80052 12.79549 12.81740 12.72858 12.79549 12.85340 12.69304
[241] 12.62951 12.76703 12.82264 12.62314 12.85646 12.86827 12.76106 12.77926 12.87170 12.74369
[251] 12.88061 12.71964 12.86959 12.90744 12.73415 12.64023 12.73952 12.62048 12.70682 12.95298
[261] 12.86007 13.00254 12.74521 13.01341 12.75380 12.71940 12.81416 12.78305 12.82399 12.93808
[271] 12.80198 12.83411 12.84774 12.89580 12.80955 12.80542 12.82215 12.81534 12.74713 12.79965
[281] 12.85922 12.75318 12.73355 12.90890 12.74630 12.76793 12.62295 12.72849 12.72374 12.62190
[291] 12.86711 12.81803 12.99207 12.88887 12.64997 12.75159 12.91875 12.87462 12.72225 12.79959
[301] 12.78466 12.77903 12.92841 12.70531 12.88677 12.71713 12.79663 12.78413 12.86081 12.89845
[311] 12.79636 12.90913 12.79055 12.88441 12.82628 12.76368 12.88808 12.83747 12.79913 12.91584
[321] 12.68815 12.83797 12.94323 12.79105 12.70656 12.73662 12.69908 12.76928 12.75728 12.69683
```



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