PROBABILITY AND STATISTICS (UCS410)
EXPERIMENT 9 :
SAMPLE STATITICS AND TESTING HYPOTHESIS

Sachin Goyal

2C42

102303557

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.

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

(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?

Now perform the following tasks:

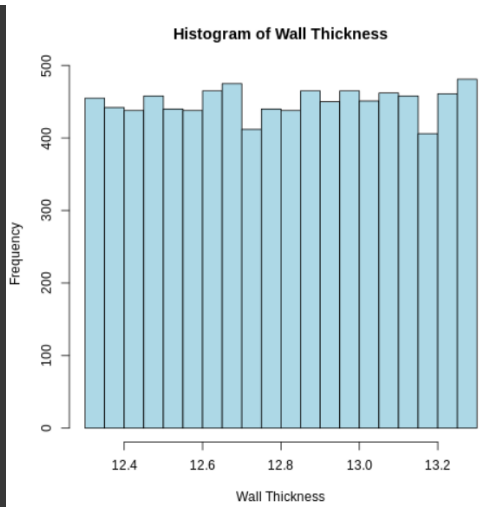
(a) Draw sufficient samples of size 10, calculate their means, and plot them in R by making histogram. Do you get a normal distribution.

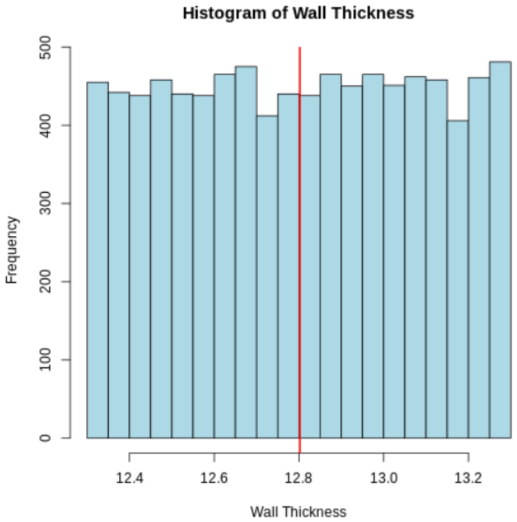
(b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.

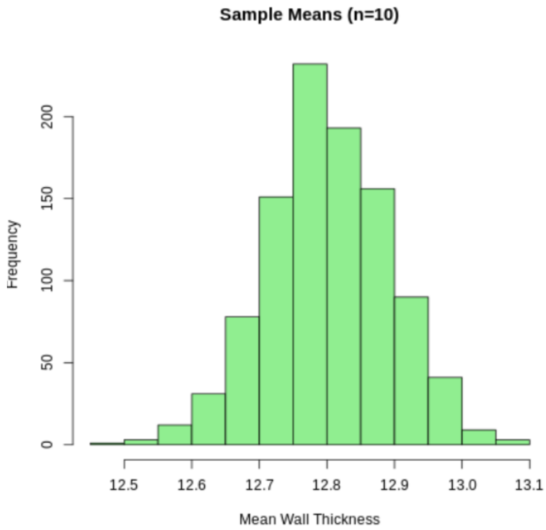
Here, we get a good bell-shaped curve and the sampling distribution approaches normal distribution as the sample sizes increase. Therefore, we can recommend the organization to use sampling distributions of mean for further analysis.

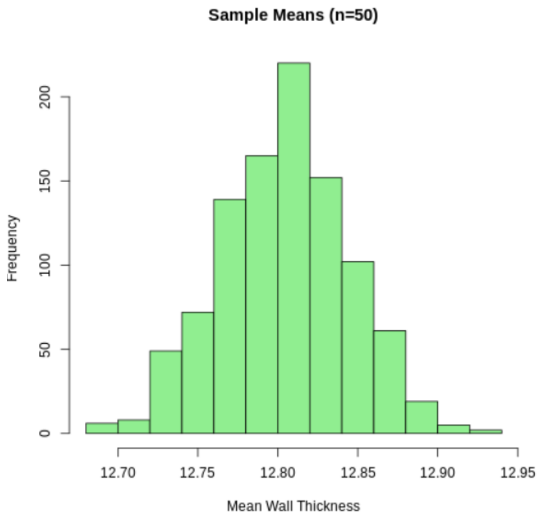
```r
# Question 1 (a)
data <- read.csv("C:/Users/Sachin Goyal/Desktop/Clt-data.csv")

# 1 (b)
nrow(data)
head(data, 10)

# 1 (c)
mean_val <- mean(data$Thickness)
hist(data$Thickness, main="Histogram of Pipe Thickness", xlab="Thickness", col="skyblue")

# (d)
abline(v=mean_val, col="red", lwd=2)

# 1 (a) Sampling Distribution (n = 10)
sample_means_10 <- replicate(1000, mean(sample(data$Thickness, 10, replace=TRUE)))
hist(sample_means_10, main="Sample Means (n=10)", xlab="Mean Thickness", col="lightgreen")

# 1 (b) Sampling Distribution (n = 50)
sample_means_50 <- replicate(1000, mean(sample(data$Thickness, 50, replace=TRUE)))
hist(sample_means_50, main="Sample Means (n=50)", xlab="Mean Thickness", col="lightblue")

# 1 (b) Sampling Distribution (n = 500)
sample_means_500 <- replicate(1000, mean(sample(data$Thickness, 500, replace=TRUE)))
hist(sample_means_500, main="Sample Means (n=500)", xlab="Mean Thickness", col="lightcoral")

# 1 (b) Sampling Distribution (n = 9000)
sample_means_9000 <- replicate(1000, mean(sample(data$Thickness, 9000, replace=TRUE)))
hist(sample_means_9000, main="Sample Means (n=9000)", xlab="Mean Thickness", col="plum")
```

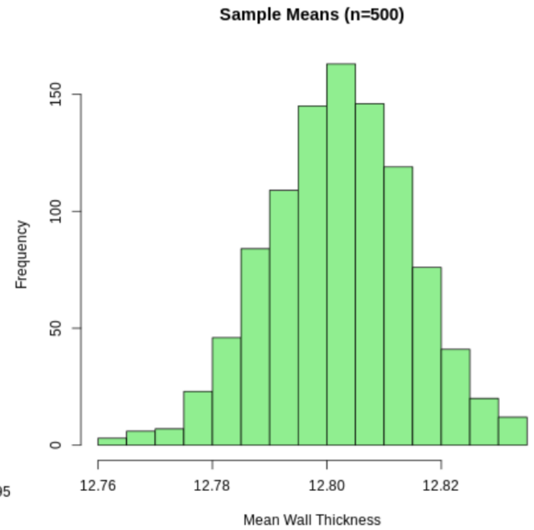
	Wall.Thickness
1	12.35487
2	12.61742
3	12.36972
4	13.22335
5	13.15919
6	12.67549
7	12.36131
8	12.44468
9	12.62977
10	12.90381

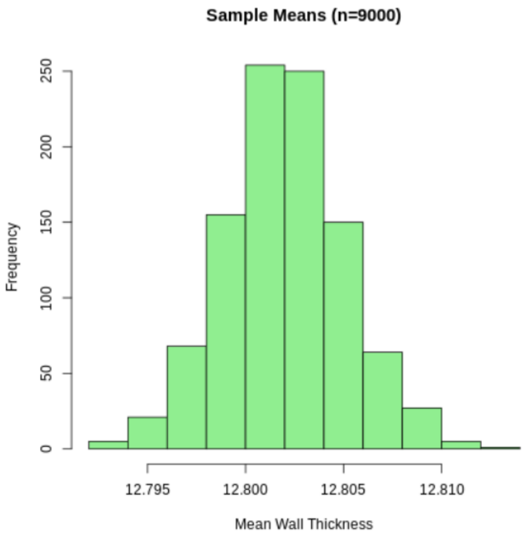



Histogram of Wall Thickness
Frequency
0
100
200
300
400
500
12.4
12.6
12.8
13.0
13.2
Wall Thickness


Sample Means (n=10)
Frequency
0
50
100
150
200
12.5
12.6
12.7
12.8
12.9
13.0
13.1
Mean Wall Thickness


Sample Means (n=50)
Frequency
0
50
100
150
200
12.70
12.75
12.80
12.85
12.90
12.95
Mean Wall Thickness


Sample Means (n=500)
Frequency
0
50
100
150
12.76
12.78
12.80
12.82
Mean Wall Thickness


Sample Means (n=9000)
Frequency
0
50
100
150
200
250
12.795
12.800
12.805
12.810
Mean Wall Thickness

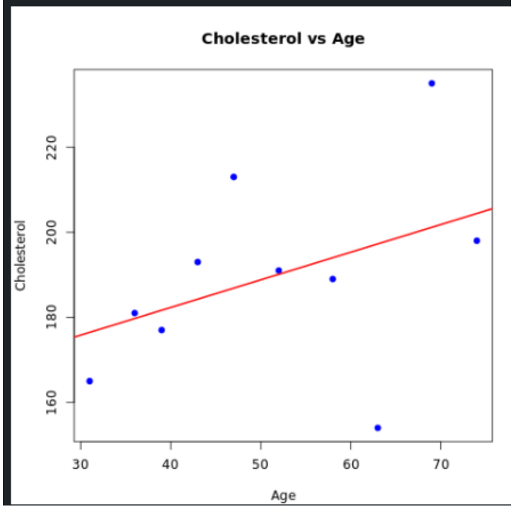
2. The following table gives information on ages and cholesterol levels for a random sample of 10 men

Age	58	69	43	39	63	52	47	31	74	36
Cholesterol	189	235	193	177	154	191	213	165	198	181

Plot the scatter diagram and a regression line that will enable us to predict Cholesterol level on age. Further, estimate the cholesterol level of a 60 year-old man.

```r

age <- c(58, 69, 43, 39, 63, 52, 47, 31, 74, 36)
cholesterol <- c(189, 235, 193, 177, 154, 191, 213, 165, 198, 181)

plot(age, cholesterol, main="Cholesterol vs Age", xlab="Age", ylab="Cholesterol", pch=19, col="blue")
model <- lm(cholesterol ~ age)
abline(model, col="red", lwd=2)

predict(model, data.frame(age = 60))

```

```
       1
195.3184

[Execution complete with exit code 0]
```



3. A research methodology course has recently been added to the PhD curriculum at the Thapar Institute of Engineering and Technology, Patiala. To evaluate its effectiveness, students take a test on formulating research problems and writing research papers both before and after completing the course. Below are the marks for a random sample of ten students:

Before the test	145	173	158	141	167	159	154	167	145	153
After the test	155	167	156	149	168	162	158	169	157	161

Assume that the differences between the pre-course and post-course test scores are normally distributed, and a high score on the test indicates a strong level of assertiveness. Do the collected data, at 5% level of significance, provide enough evidence to conclude that research scholars become more assertive after completing the course?

```
before <- c(145, 173, 158, 141, 167, 159, 154, 167, 145, 153)
after <- c(155, 167, 156, 149, 168, 162, 158, 169, 157, 161)
t.test(after, before, paired=TRUE, alternative="greater")
```

```
	Paired t-test

data:  after and before
t = 2.2597, df = 9, p-value = 0.0251
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.7551657       Inf
sample estimates:
mean of the differences 
                      4 
```