

Probability and Statistics (UCS410)

Experiment 8

Name: Ketan Singh

Group: 3CO14

Roll no.: 102003349

- 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.

```
#(a) importing dataset
data = read.csv(file.choose())

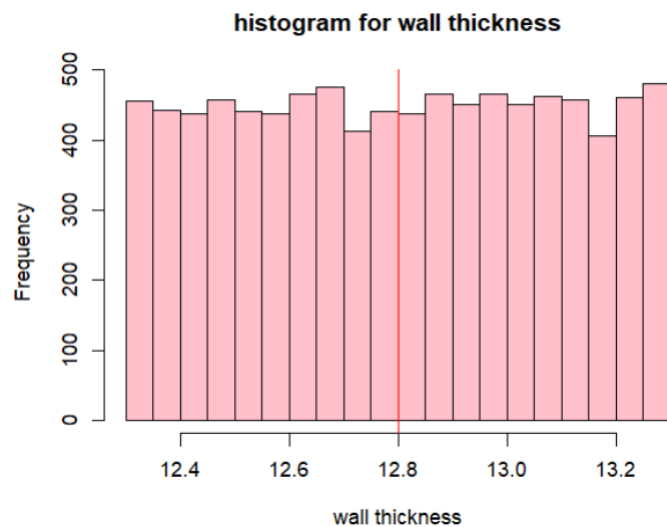
#(b) count of rows and columns
dim(data)
#view top 10 rows of dataset
head(data,10)

#(c) calculate the population mean
mean(data$wall.Thickness)
#plot all the observations in the data
hist(data$wall.Thickness,col="pink",main="histogram for wall thickness",xlab="wall thickness")

#(d) adding straight lines to a plot
abline(v=12.8,col="red",lty=1)
```

Output:

```
> #(a) importing dataset
> data = read.csv(file.choose())
> #(b) count of rows and columns
> dim(data)
[1] 9000 1
> #view top 10 rows of dataset
> 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
> mean(data$wall.Thickness)
[1] 12.80205
> #plot all the observations in the data
> hist(data$wall.Thickness,col="pink",main="histogram for wall thickness",xlab="wall thickness")
Error in plot.new() : figure margins too large
> #(d) adding straight lines to a plot
> abline(v=12.8,col="red",lty=1)
Error in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...) :
  invalid graphics state
```



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.

```
#(a)for sample size=10 ,samples=9000
#calculate the arithmetic mean and plot the mean of sample 9000 times
s10<-c()
n=9000
for(i in 1:n){
  s10[i] = mean(sample(data$Wall.Thickness,10,replace=TRUE))
}
hist(s10,col="lightgreen",main="Sample size = 10",xlab="wall thickness")
abline(v = mean(s10),col="blue")
abline(v = 12.8, col="blue")

#(b)for sample size=50,500,9000
s50<-c()
s500<-c()
s9000<-c()
n=9000
for(i in 1:n){
  s50[i] = mean(sample(data$Wall.Thickness,50,replace=TRUE))
  s500[i] = mean(sample(data$Wall.Thickness,500,replace=TRUE))
  s9000[i] = mean(sample(data$Wall.Thickness,9000,replace=TRUE))
}
par(mfrow=c(1,3))
hist(s50,col="lightblue",main="Sample size = 50",xlab = "wall thickness")
abline(v=mean(s50), col="red")

hist(s500,col="lightblue",main="Sample size = 500",xlab = "wall thickness")
abline(v=mean(s500), col="red")

hist(s9000,col="lightblue",main="Sample size = 9000",xlab = "wall thickness")
abline(v=mean(s9000), col="red")
```

Output:

