

Probability and Statistics (UCS410)

Experiment 8

- 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:

- Import the csv data file in R. `store data like this data<-read.csv(file.choose())`
- Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset. `dim(data)`, or `nrow(iris)`, or `ncol(iris)`, and use `head(data)` to print first 6 rows and `head(data,10)` for first 10 rows
- Calculate the population mean and plot the observations by making a histogram. `hist(data$Wall.Thickness)` `mean(data$Wall.Thickness)`
- Mark the mean computed in last step by using the function `abline`. `abline(v=meanLol,col="red")`

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

Now perform the following tasks:

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