Central Limit Theorm

Yash Singhal

August 8, 2018

Overview

This Work attemps to show the Central Limit Theorm in pratice via Simulation using R language

The central limit theorem states that even if a population distribution is strongly non-normal, its sampling distribution of means will be approximately normal for large sample sizes (over 30). Hence it is widely Used as it allows the stasticians to work . with the distributions that are not inherently normal.

Hence in the next couple of slides we will go over the consequneces of central limit theorm particularly the mean and standard deviation

Simulations

```
Sample <- rexp(n=1000,rate = 0.2)
```

Here I have taken a sample of 1000 observations from random exponential distribution and stored in the variable sample

```
Sampling_Distribution <- NULL
for (i in 1:40){
  Sampling_Distribution <-c(Sampling_Distribution,mean(rexp(n=1000,rate=0.2)))
}
```

The following code stores the 40 sample means where each sample has a size of 1000 observations. The resultant distribution of these 40 samples constituents what is called a Sampling Distribution

Normal Distrbution

Here we will go over the evidence of Normal Distribution which is stated in CLT

```
paste("Mean of Sample",mean(Sample))
```

```
## [1] "Mean of Sample 4.96002892535145"
```

The sample is highly skewed with the right tail pushing the mean further out from Zero(Figure 1)

```
paste("Mean of Sampling distribution",mean(Sampling_Distribution))
```

```
## [1] "Mean of Sampling distribution 5.00884276712855"
```

The sampling distribution starts to look more like normal here with accordance of the Central Limit theorm. The mean is also nearly situated around middle(Figure 2)

It is noted that Central Limit also states that the mean of sample means closely resembles the population average

Similar Situation can also be seen in the case of standard deviation

```
sd(Sample)
```

```
## [1] 4.803043
```

The standard deviation of the Sample is quite large which is expected owning to Highly skewed nature(Figure 1)

```
sd(Sampling_Distribution)
```

```
## [1] 0.1418524
```

Here the Values are highly Concentrated due to normal form of the sampling distrubtion (Figure 2)

If the Sampling distribution is indeed normal the distrubution should be a straight line in qqplot . Here the sampling distribution closely resembles normal with a few missteps only(Figure 3) . With more samples evidently it will start to look more like a normal distribution

Theoritical Mean and Therotical Std vs Sampling Stastics

According to Central Limit theorm the population parmater closely resembles the average of the samples in sampling distrbution while sampling standard distribution is more closely resembles the standard distrubution of the population over square root of no

Since here our population parameter is absent our sample stastic would be next better guess for mean and pollution

```
mean(Sample) - mean(Sampling_Distribution)
```

```
## [1] -0.04881384
```

```
sd(Sample/sqrt(1000))-(sd(Sampling_Distribution))
```

```
## [1] 0.01003319
```

Both the above closely limit to zero as n approaches Infinity (sample size) as accordance with CLT

Appendix

```
hist(Sample,main="Sample",xlab="Output Value")
```

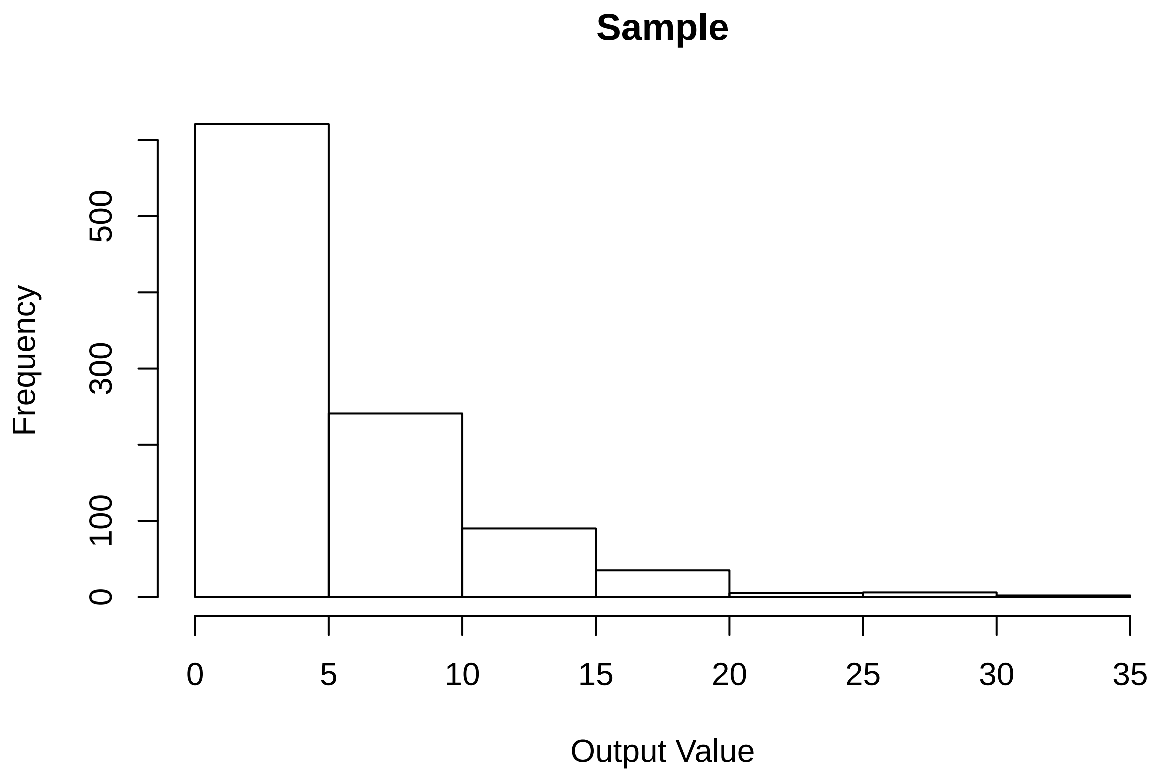

Figure 1 : Histogram of the Sample Observations

```
hist(Sampling_Distribution,main="Sampling Distribution",xlab="Output")
```

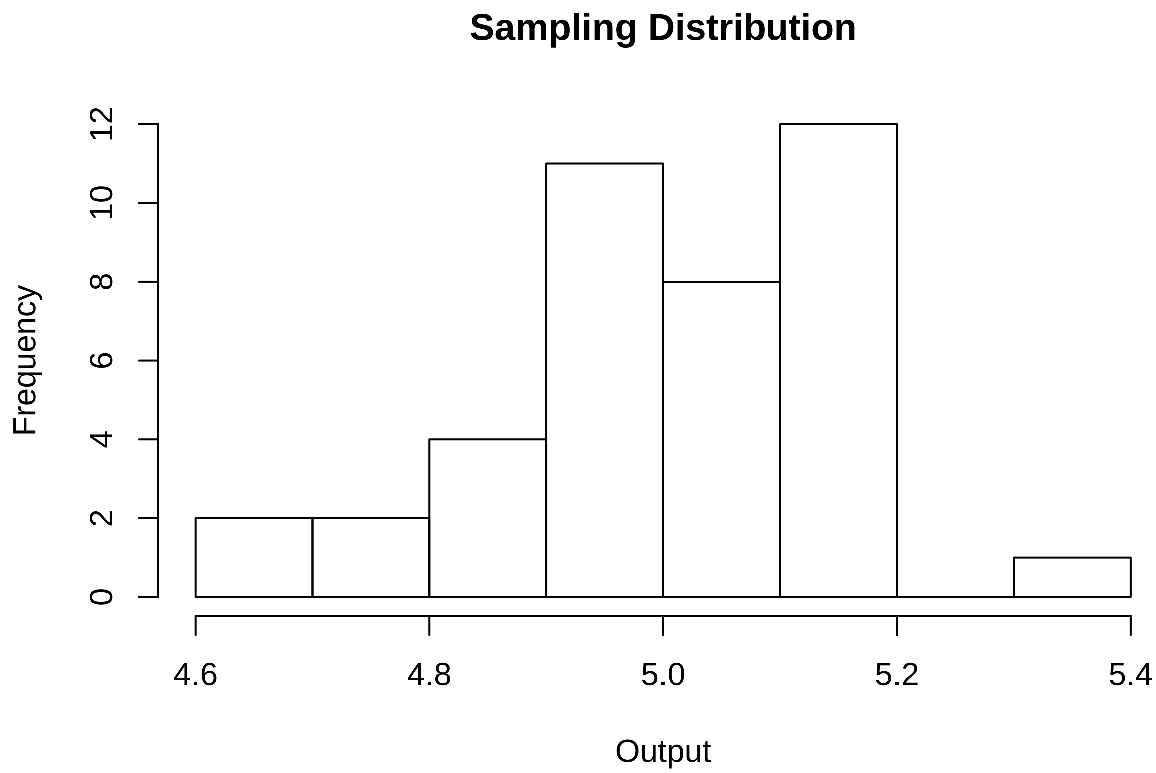

Figure 2: Histogram of Sampling Distribution

```
qqnorm(Sampling_Distribution)
```

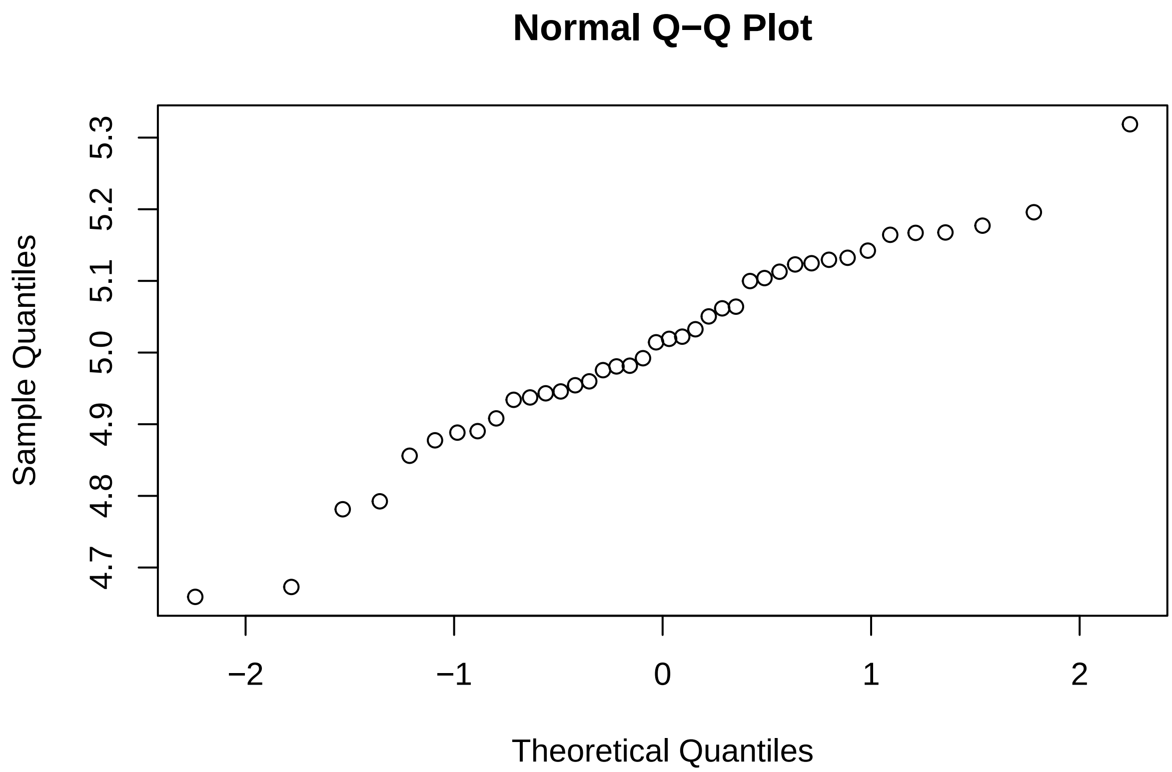

Figure 3 : Normal quantle plot