

title: "Statistical Inference Course Project Part 1"

author: "dinh quy kien"

date: "May 09, 2023"

output:

pdf\_document: default

html\_document: default

word\_document: default

Synopsis The project consists of two parts:

1. Simulation Exercise to explore inference
2. Basic inferential analysis using the ToothGrowth data in the R datasets package

Part 1: Simulation Exercise

The task is to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution will be simulated in R with `rexp(n,lambda)` where `lambda` is the rate parameter. The mean of exponential distribution and the standard deviation are both  $1/\lambda$  where  $\lambda = 0.2$ , and distribution of averages of 40 exponentials and will perform 1000 simulations.

### Mean Comparision Sample Mean vs Theoretical Mean of the Distribution

```
# Sample Mean
sampleMean <- mean(mean_sim_data) # Mean of sample means
print (paste("Sample Mean = ", sampleMean))

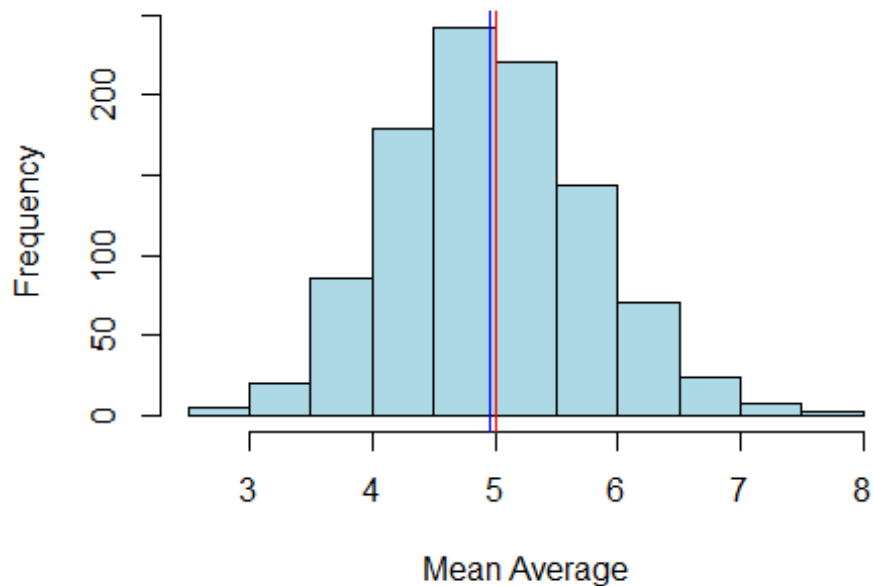
## [1] "Sample Mean = 4.96640993016276"

# Theoretical Mean
# the expected mean of the exponential distribution of rate = 1/Lambda
theoretical_mean <- (1/lambda)
print (paste("Theoretical Mean = ", theoretical_mean))

## [1] "Theoretical Mean = 5"

# Histogram shows differences
hist(mean_sim_data, col="light blue", xlab = "Mean Average", main="Distributi
on of Exponential Average")
abline(v = theoretical_mean, col="red")
abline(v = sampleMean, col="blue")
```

## Distribution of Exponential Average



**Question 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution**

Calculating the theoretical and sample variance

```
# sample deviation & variance
sample_dev <- sd(mean_sim_data)
sample_dev

## [1] 0.802972

sample_variance <- sample_dev^2
sample_variance

## [1] 0.644764

# theoretical deviation & variance
theoretical_dev <- (1/lambda)/sqrt(n)
theoretical_dev

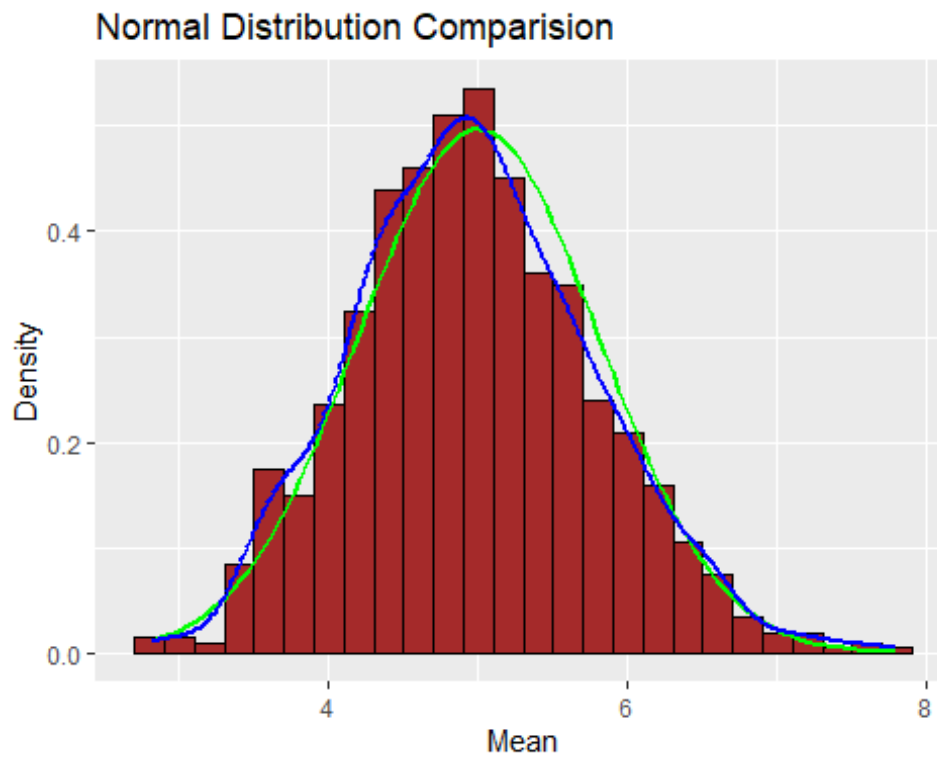
## [1] 0.7905694

theoretical_variance <- ((1/lambda)*(1/sqrt(n)))^2
theoretical_variance

## [1] 0.625
```

**Question 3: Show that the distribution is approximately normal Histogram with Density and sample means:**

```
d <- data.frame(mean_sim_data)
t <- data.frame(theoretical_mean)
g <- ggplot(d, aes(x = mean_sim_data)) + geom_histogram(binwidth = .2, color=
"black", fill="brown" , aes(y=..density..))+stat_function(fun=dnorm, args=li
st(mean=theoretical_mean, sd=sd(mean_sim_data)), color="green", size =1) + st
at_density(geom = "line", color = "blue", size =1) + labs(x="Mean", y= "Dens
ity", title="Normal Distribution Comparision")
```



The above plot indicated that density curve is similar to normal distribution curve.

Q-Q Normal Plot also indicates the normal distribution

```
qqnorm(mean_sim_data) qqline(mean_sim_data, col = "magenta")
```

Normal Q-Q Plot

