# Peer-graded Assignment: Statistical Inference Course Project

Part 1: Simulation Exercise Instructions

Sharanya Gowda

July 29, 2022

## 1. PROBLEM STATEMENT

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with `rexp(n, lambda)` where lambda is the rate parameter. The **mean of exponential distribution is 1/lambda** and the **standard deviation is also 1/lambda**. Set **lambda = 0.2** for all of the simulations. You will **investigate the distribution of averages of 40 exponentials**. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should:

- Show the sample mean and compare it to the theoretical mean of the distribution.
- Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- Show that the distribution is approximately normal.

*In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.*

## 2. SOLUTION

- Number of exponentials (n) = 40
- Lambda (l) = 0.2 iterations
- Simulations (sim) = 1:1000
- Mean of exponential distribution (mean)= 1/l = 1/0.2 = 5
- Standard Devialtion (std) = 1/l = 1/0.2 = 5

```
set.seed(1)
n <- 40
l <- 0.2
sim <- 1:1000
```

**First, generate a data set.** The exponential distribution values are simulated using the `rexp` function:

```
Exp_Dist_Data <- data.frame(values = sapply(sim, function(values) {
         mean(rexp(n, l))
        }))

head(Exp_Dist_Data,5)
```

```
##     values
## 1 4.860372
## 2 5.961285
## 3 4.279204
## 4 4.702298
## 5 5.196446
```

### 2.1. Sample Mean vs.Theoretical Mean

```
SMean_values<- apply(Exp_Dist_Data, 2, mean)
SMean_values
```

```
##   values
## 4.990025
```

```
TMean_values <- 1/l
TMean_values
```

```
## [1] 5
```

- **RESULT - The Sample Mean is almost equal to the Theoretical Mean.(())

### 2.2. Sample Variance vs.Theoretical Variance

```
STD_values <- sd(Exp_Dist_Data$values)
STD_values
```

```
## [1] 0.7817394
```

```
TSTD_values <- (1/l)/sqrt(n)
TSTD_values
```

```
## [1] 0.7905694
```

```
Variance_values <- var(Exp_Dist_Data$values)
Variance_values
```

```
## [1] 0.6111165
```

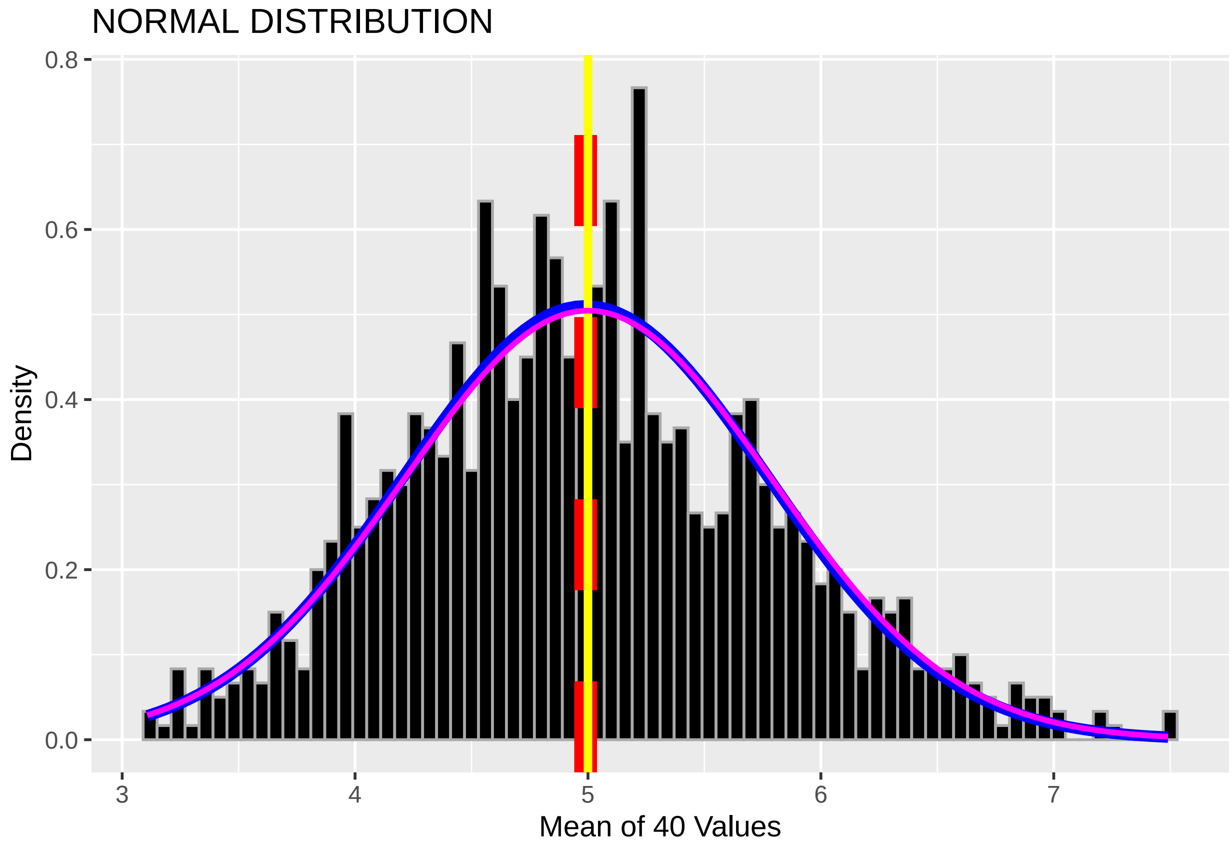
```
TVariance_values <- TSTD_values^2
TVariance_values
```

```
## [1] 0.625
```

- **RESULT - The Sample Variance is almost equal to the Theoretical Variance.**

### 2.3. Normal Distribution

```
library(ggplot2)
Nor_Dist <- ggplot(data = Exp_Dist_Data, aes(x = values)) + geom_histogram(aes(y=..density..), binwidth
Nor_Dist <- Nor_Dist + labs(title="NORMAL DISTRIBUTION", x="Mean of 40 Values", y="Density")
Nor_Dist <- Nor_Dist + geom_vline(xintercept=SMean_values,size=4, color="red", linetype=2)
Nor_Dist <- Nor_Dist + stat_function(fun=dnorm,args=list(mean=SMean_values, sd=STD_values),color = "blu
Nor_Dist <- Nor_Dist+ geom_vline(xintercept=TMean_values,size=1.5,color="yellow",linetype = 1)
Nor_Dist <- Nor_Dist + stat_function(fun=dnorm,args=list(mean=TMean_values, sd=TSTD_values),color = "ma
Nor_Dist
```




- **RESULT - The output plot shows that the distribution is almost normal.**