# Statistical Inference Course Project Part1

Khalid

7/11/2020

# Part ONE

## The project consists of two parts:

- **A simulation exercise**.
- Basic inferential data analysis.

### 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 averag
es 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 t
he distribution.
- Show that the distribution is approximately normal.
```

### install needed packages

```
library(cowplot)
```

```
##
## ********************************************************
```

```
## Note: As of version 1.0.0, cowplot does not change the
```

```
##   default ggplot2 theme anymore. To recover the previous
```

```
##   behavior, execute:
##   theme_set(theme_cowplot())
```

```
## ********************************************************
```

```
library(ggplot2)
```

# assigning values

```
lambda <- 0.2
sample_num <- 40
sim_num <- 1000
```

# calculate the 40 samples 1000 times

```
set.seed(2) #for reproducibility purposes
mean_sim_num <- replicate(1000, mean(rexp(sample_num, rate = lambda)))
```

# Show the sample mean and compare it to the theoretical mean of the distribution.

# The mean of exponential distribution is 1/lambda

```
1/lambda
```

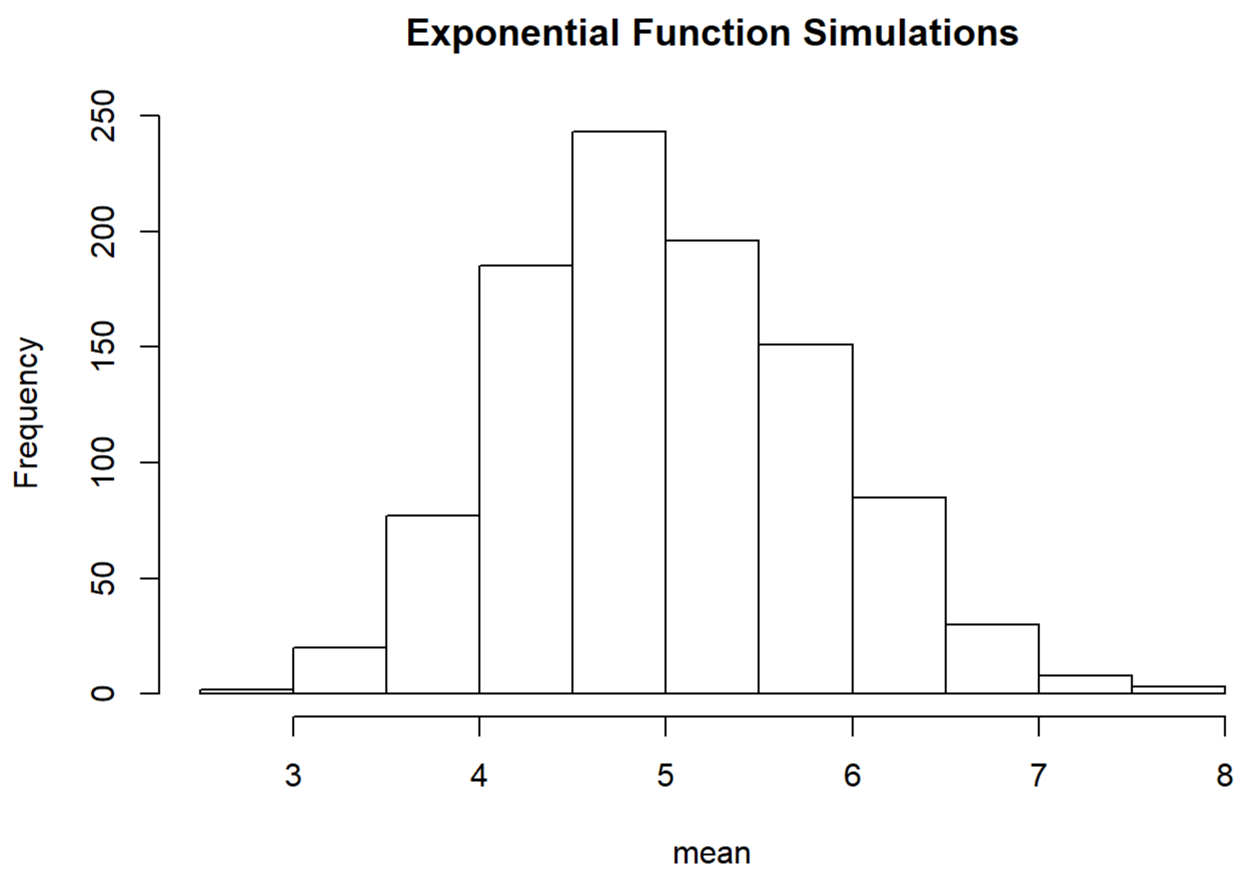
```
## [1] 5
```

```
# The mean othe 40 samples 1000 times
mean(mean_sim_num)
```

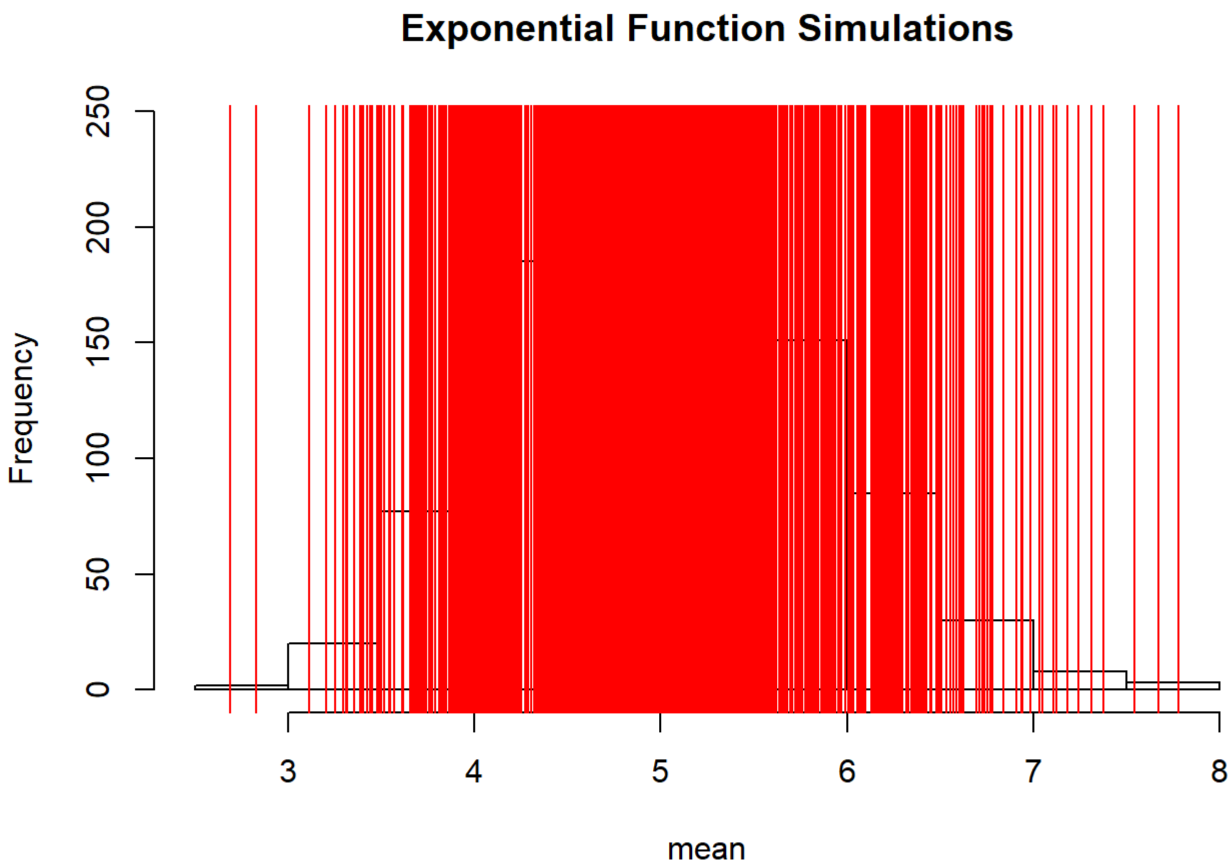
```
## [1] 5.016356
```

# visualization

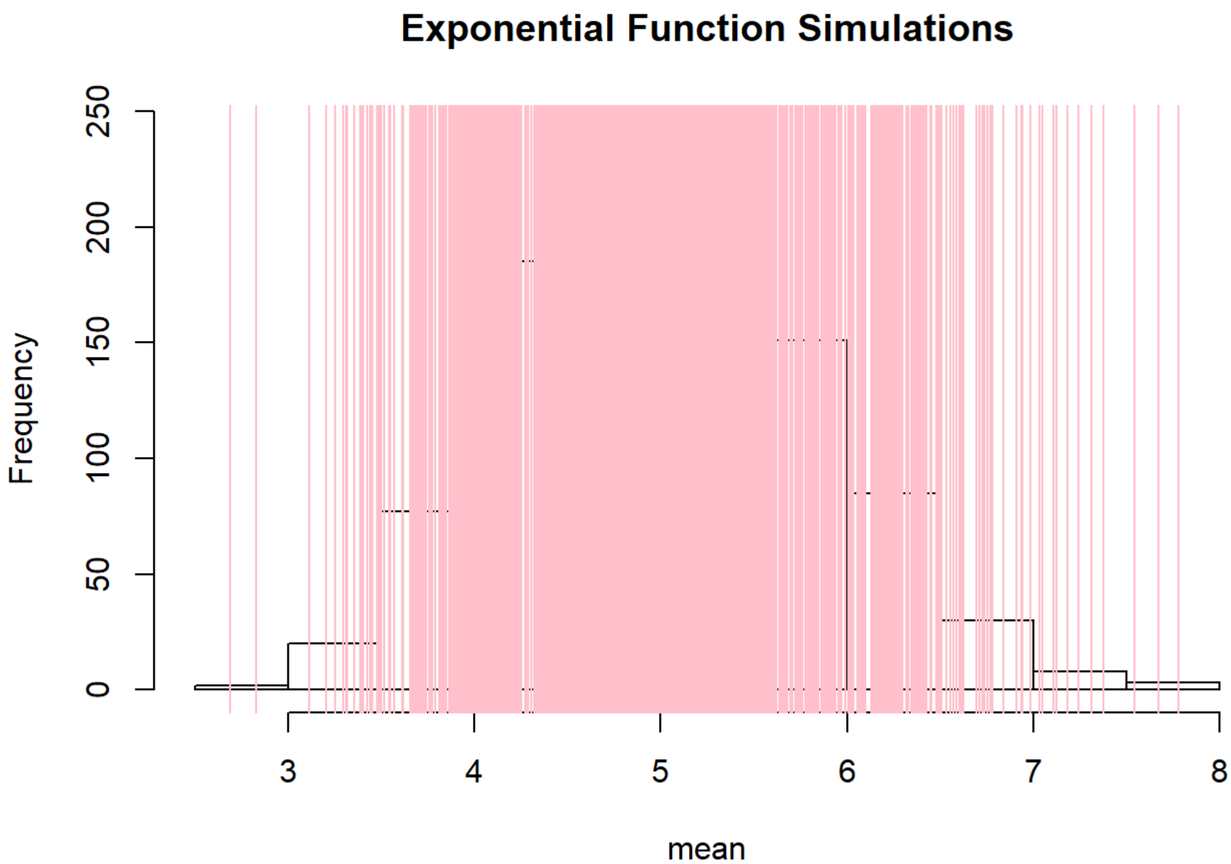
```
hist(mean_sim_num, xlab = "mean", main = "Exponential Function Simulations")
```



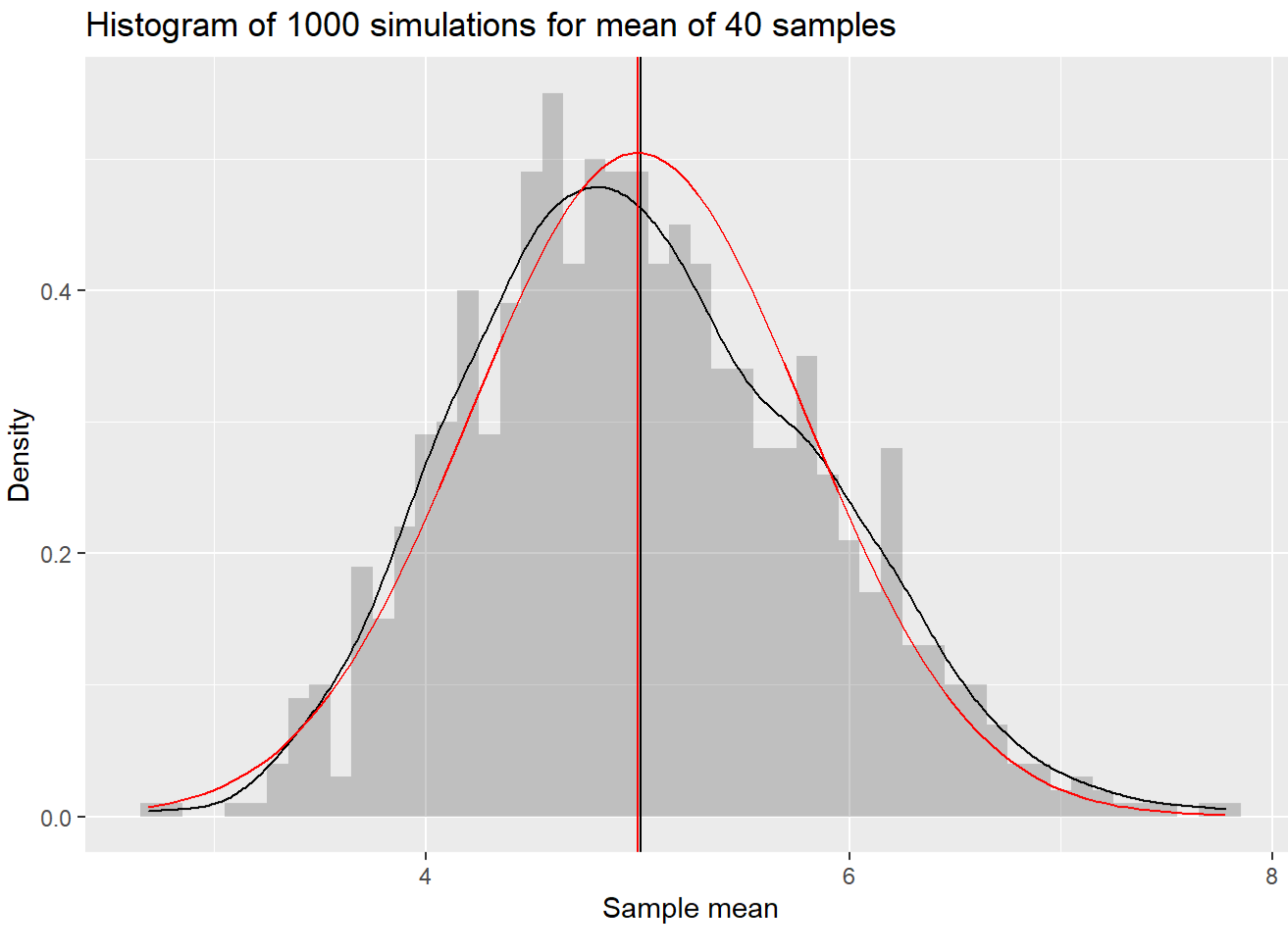

```
hist(mean_sim_num, xlab = "mean", main = "Exponential Function Simulations")
abline(v = mean_sim_num, col = "red")
```




```
hist(mean_sim_num, xlab = "mean", main = "Exponential Function Simulations")
abline(v = mean_sim_num, col = "pink")
```




```
ggplot(data = as.data.frame(mean_sim_num), aes(x = mean_sim_num)) +
        geom_histogram(binwidth = 0.1, aes(y = ..density..), alpha = 0.3) +
        geom_density(color = 'black') + #density curve of the sample distribution
        geom_vline(xintercept = mean(mean_sim_num), color = 'black') +
        stat_function(fun = dnorm, args = list(mean = 1/lambda, sd = 1/lambda/sqrt(sample_num)),
color = 'red') +
        geom_vline(xintercept = 1/lambda, color = 'red') +
        xlab('Sample mean') +
        ylab('Density') +
        ggtitle('Histogram of 1000 simulations for mean of 40 samples')
```




# Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

### standard deviation of distribution

```
sd_dv_dist <- sd(mean_sim_num)
sd_dv_dist
```

```
## [1] 0.818004
```

### standard deviation from analytical expression

```
sd_dv_theory <- (1  / lambda) / sqrt(sample_num)
sd_dv_theory
```

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

### variance of distribution

```
variance_dist <- sd_dv_dist^2
variance_dist
```

```
## [1] 0.6691305
```

### variance from analytical expression

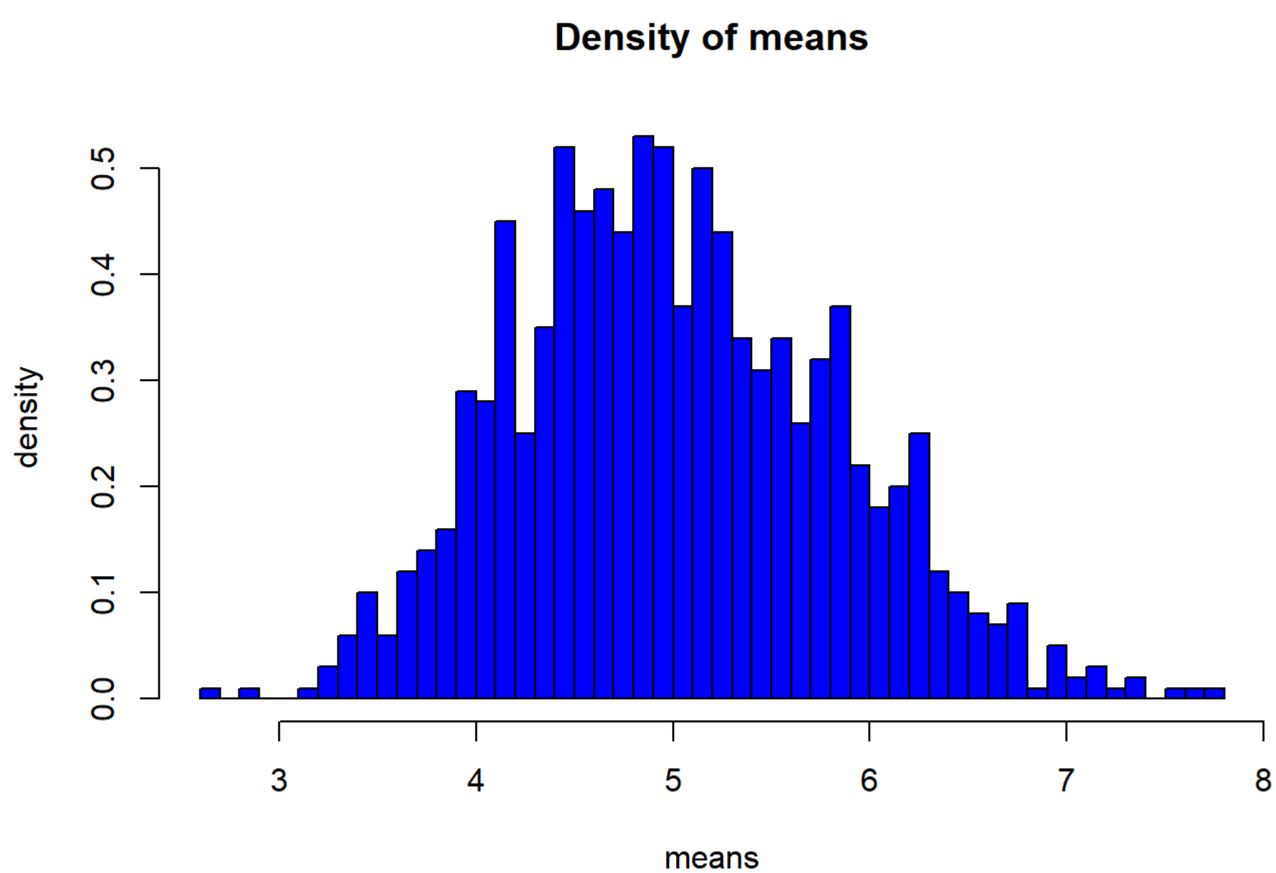
```
variance_theory <- ((1 / lambda)*(1 / sqrt(sample_num)))^2
variance_theory
```

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

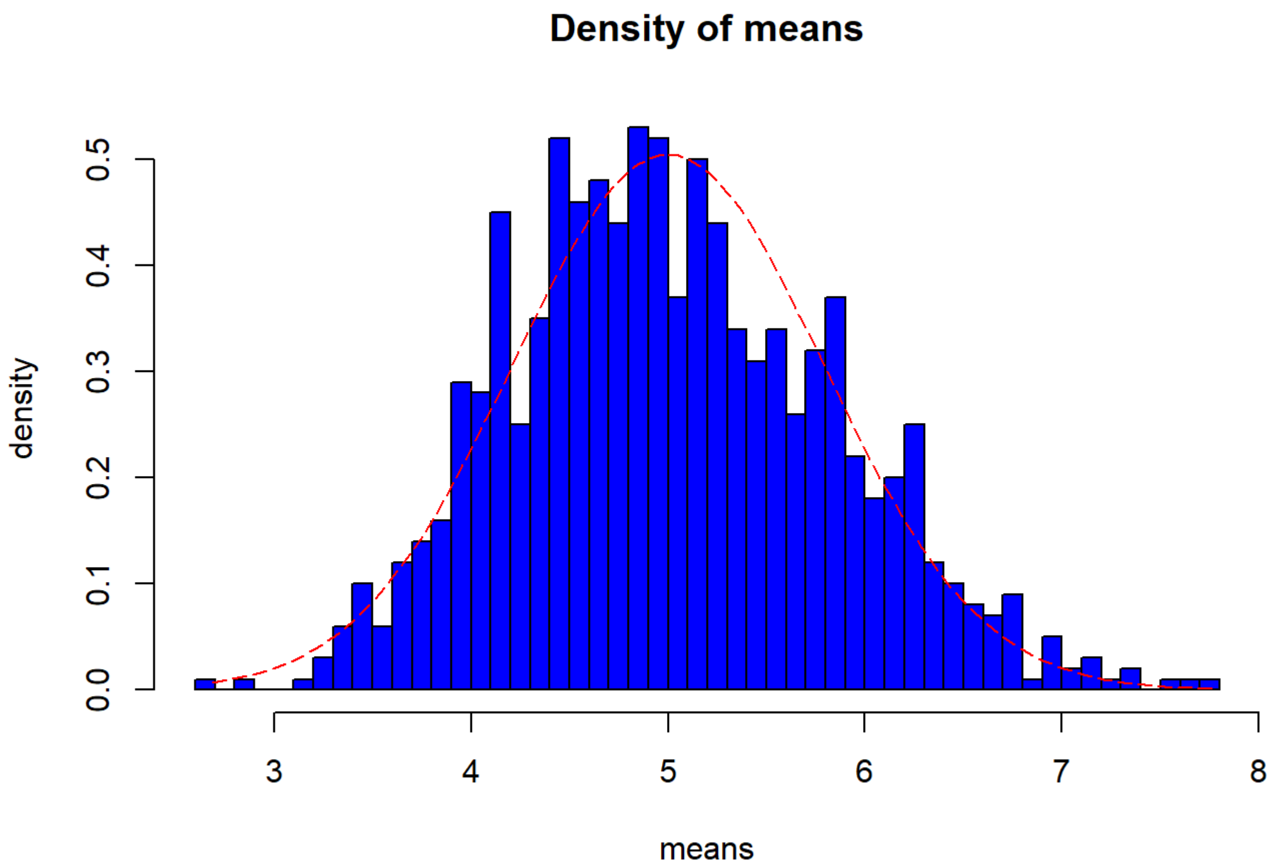
### Show that the distribution is approximately normal.

```
xfit <- seq(min(mean_sim_num), max(mean_sim_num), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_num)))
```

```
hist(mean_sim_num,breaks=sample_num,prob=T,col="blue",xlab = "means",main="Density of means",yla
b="density")
```

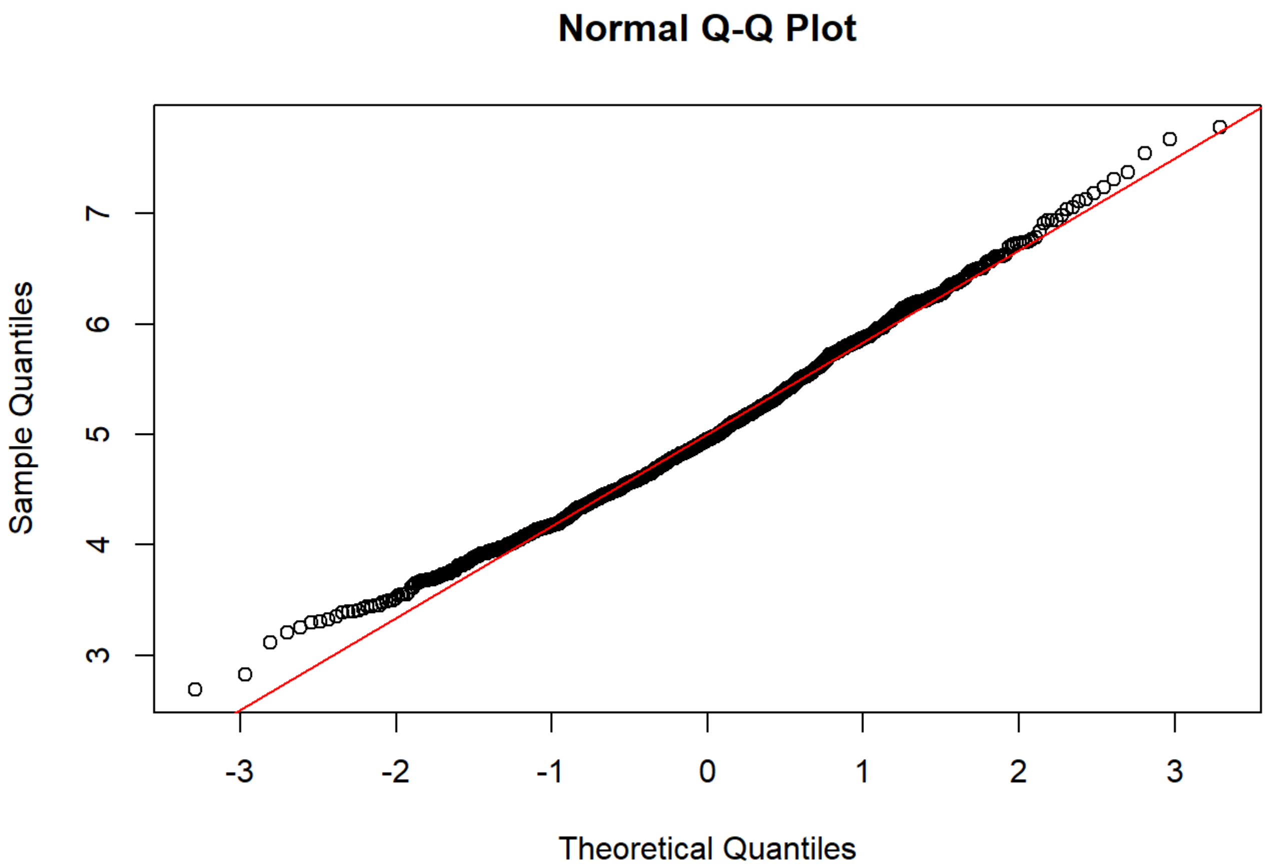



```
hist(mean_sim_num,breaks=sample_num,prob=T,col="blue",xlab = "means",main="Density of means",yla
b="density")
lines(xfit, yfit, pch=22, col="red", lty=5)
```




## compare the distribution of averages of 40 exponentials to a normal distribution

```
qqnorm(mean_sim_num)
qqline(mean_sim_num, col = 2)
```


Normal Q-Q Plot
Sample Quantiles
3
4
5
6
7
-3
-2
-1
0
1
2
3
Theoretical Quantiles