# Statiscal Inference Course Project Part1

Andres Felipe Lagos Ruiz

9/18/2020

In this project it is going to investigate the exponential in R and compare it with the Central Limit Theorem (CLM). The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda. using lambda = 0.2 for all of the simulations. It will investigate the distribution of averages of 40.
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential.
1. Show the sample mean and compare it to the theoretical mean of the distribution
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution
3. Show that the distribution is approximately normal

```
library(ggplot2)
```

### Build the data

```
#set the randomness
set.seed(1020)
#Inicialze the variables
lambda<-0.2
n<-40
experiments<-1000
```

### make the experiments

```
#make the data
samples<-matrix(rexp(n*experiments,rate=lambda),experiments)
#aplly the mean to the sample data
hist<-data.frame(apply(samples, 1, mean))
```

### Point 1: Show the sample mean and compare it to the theoretical mean of the distribution

```
#sample mean
sm<-mean(hist$apply.samples..1..mean.)
sm
```

```
## [1] 5.026244
```

```
#theorical mean
tm<-1/lambda
tm
```

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

```
#diference between the means
sm-tm
```

```
## [1] 0.02624403
```

It is possible to say that the differences between the mean is small

### Point 2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
#sample variance
sv<-var(hist$apply.samples..1..mean.)
sv
```

```
## [1] 0.6175483
```

```
#theorical variance
tv<-(1/lambda)^2/ n
tv
```

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

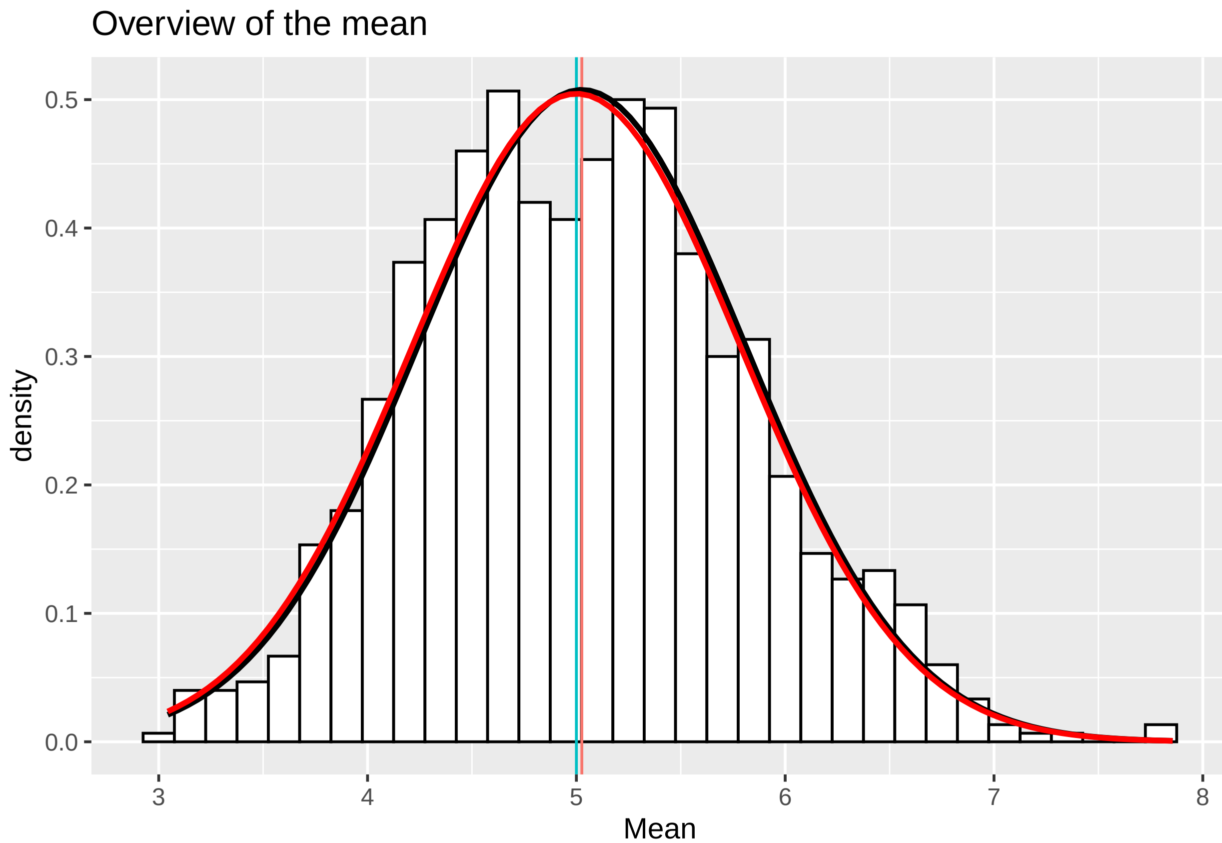
```
#diference between the variances
sv-tv
```

```
## [1] -0.00745174
```

The results are similar with a difference of 0.007.

### 3 Show that the distribution is approximately normal

```
g=ggplot(hist, aes(x=apply.samples..1..mean.))
gh=g+geom_histogram(binwidth = 0.15,color="black",fill="white",aes(y=..density..))
gh<-gh+xlab("Mean")+labs(title = "Overview of the mean")
gh<- gh + geom_vline(aes(xintercept = sm, colour = "green"))
gh <- gh + geom_vline(aes(xintercept = tm, colour = "yellow"))
gh<-gh + stat_function(fun = dnorm, args = list(mean = sm, sd = sqrt(sv)), color = "black", size = 1.0)
gh <- gh + stat_function(fun = dnorm, args = list(mean = tm, sd = sqrt(tv)), colour = "red", size = 1.0
gh+theme(legend.position="none")
```




seen the graphic it is possible to say that is not exactly a normal distributio but it is similar. Other way to make this is with the Ci of the theoretical normal distribution and the actual answers of the distribution.

```
#CI for the sample distribution
round (mean(sm) + c(-1,1)*1.96*sqrt(sv)/sqrt(n),3)
```

```
## [1] 4.783 5.270
```

```
#CI for the theorical distribution
 round (mean(tm) + c(-1,1)*1.96*sqrt(tv)/sqrt(n),3)
```

```
## [1] 4.755 5.245
```

the results of the CI of 0.95 shows the sames that the histogram

## Conclucion

it is possible to say that this agree with CLT