```
Author: Giuseppa Cefalu
Date: 9/29/2020

rm(list=ls())

library(ggplot2)
library(gridExtra)
library(knitr)
library(rmarkdown)


#THE EXPONENTIAL DISTRIBUTION AND THE CENTRAL LIMIT THEOREM

#his projects applies the Central Limit Theorem to 1000 and 10000 exponential distribution
#simulations of size 40 samples, and 1000 simulations of sample size 60 - lambda 0.2 .The
#simulated mean is commpared to the theoretical mean - 1/lambda - and the simulated
#variance is compared to the theoretical variance - 1/lambda -.The distribution of
#averages of 1000/10000 simulations of sample size 40 demonstrates that the
#disitribution of averages aproached a normal distribution as the number of
#simulations increases and the distribution of averages of 1000 samples of size 60
#demonstrates that the distribution of  averages approaches a normal distribution as the
#sample size increases as inidcated by the simulated versus normal distribution and
#Quantile -Quantile plots. Additionally, comparison of the simulated 95% confidence
#interval with the theoretical 95% confidence interval shows that the true mean is
#likely to lie within that range,in other words 95% of the samples will contain the
#mean of the population.


#Average number of events per unit of time
lambda <- 0.2
#Sample size
sample <- 40
#Number of samples
samples <- 1000
#Expected or theoretical mean
expected <- 1/lambda


#Exponential distribution - One sample (size 1000)
data <- rexp(1000, lambda)
data <- data.frame(data, 1000)

#Exponential distribution
ex <- ggplot(data, aes(x = data)) +
        geom_histogram(aes(y = ..density..), alpha = 0.2, binwidth = 0.5, col = "black") +
        ylim(c(0,0.3)) +
        ggtitle("Exponential Distribution") +
    theme(plot.title=element_text(face="bold", size=10))

#Set reproducible results for verification
set.seed(1)

#generate the distribution (1000 samples) of the averages of 40 exponentials with Lambda
```

```
#0.2
means = NULL
for (i in 1 : 1000) means = c(means, mean(rexp(sample, lambda)))
data <- data.frame(means,size=40)

#SAMPLE MEAN VERSUS THEORETICAL MEAN - 1/lambda - 1000 simulations - n = 40
#mean of means
meanOfMeans <- mean(means)
cat("Mean of means:" ,meanOfMeans)
```

```
## Mean of means: 4.990025
```

```
#theoretical mean
theoreticalMean <- (1/0.2)
cat("\nTheoretical mean:", theoreticalMean)
```

```
##
## Theoretical mean: 5
```

```
#SAMPLE VARIANCE VERSUS THEORETICAL VARIANCE - 1000 simulations - n = 40
sampleVariance <- var(means)
cat("\nSample variance:", sampleVariance)
```

```
##
## Sample variance: 0.6111165
```

```
theoreticalVariance <- ((1/0.2)/sqrt(40))^2
cat("\nTheoretical variance:", theoreticalVariance)
```

```
##
## Theoretical variance: 0.625
```

```
#Plot distribution
p <- ggplot(data, aes(x = means, fill = 40)) + theme_bw() +
     geom_histogram(aes(y = ..density..), alpha = 0.7, binwidth = 0.30, col = "black") +
     ylim(c(0,0.6)) +
     stat_function(mapping = NULL, data = NULL, geom = "path", position = "identity",
           fun=dnorm, n = 101, size = 1,  args=list(mean=5,s =sd(means))) +
    geom_vline(aes(xintercept=mean(means),colour="red"), size = 0.2) +
    geom_text(aes(x = mean(means), data = NULL, label="\nmean",y=0.2),
           colour="black",angle=90, size = 3) +
    ggtitle("Smulated distribution verus
normal distribution - 1000 simulations - sa
mple size 40") +
    theme(plot.title=element_text(face="bold", size=9))

#Quantile-quantile plot of sample variable versus theoretical
qq <- ggplot(data, aes(sample = data[,1])) +  stat_qq() + stat_qq_line()

#THE SIMULATED DISTRIBUTION APPRACHES THE NORMAL DISTRIBUTION AS THE NUMBER OF SIMULATIONS
#INCREASES
newMeans = NULL
for (i in 1 : 10000) newMeans = c(newMeans, mean(rexp(40, lambda)))
newData <- data.frame(newMeans,size=40)

#Plot distribution
newp <- ggplot(newData, aes(newMeans, fill = 30)) + theme_bw() +
        geom_histogram(aes(y = ..density..), alpha = 0.7,binwidth = 0.30,  col = "black") +
        stat_function(mapping = NULL, data = NULL, geom = "path", position = "identity",
              fun=dnorm, size = 1, args=list(mean=5,s =sd(means)))  +
        geom_vline(aes(xintercept=mean(means),colour="red")) +
        geom_text(aes(x = mean(means), data = NULL,size = 0.1, label="\nmean",y=0.2),
              colour="black",angle=90, size = 3) +
        ggtitle("Smulated distribution verus
normal distribution - 10000 simulations -
sample size 40") +
        theme(plot.title=element_text(face="bold", size=9))

#SAMPLE MEAN VERSUS THEORETICAL MEAN - 1/lambda - 10000simulations n = 40
#mean of means
meanOfMeans <- mean(newMeans)
cat("Mean of means:" ,meanOfMeans)
```

```
## Mean of means: 5.002635
```

```
#theoretical mean
theoreticalMean <- (1/0.2)
cat("\nTheoretical mean:", theoreticalMean)
```

```
##
## Theoretical mean: 5
```

```
#SAMPLE VARIANCE VERSUS THEORETICAL VARIANCE - 10000 simulations n = 40
sampleVariance2 <- var(newMeans)
cat("\nSample variance:", sampleVariance2)
```

```
##
## Sample variance: 0.6194882
```

```
theoreticalVariance2 <- ((1/0.2)/sqrt(40))^2
cat("\nTheoretical variance:", theoreticalVariance2)
```

```
##
## Theoretical variance: 0.625
```

```
#Quantile-Quantile plot
newqq <- ggplot(newData, aes(sample = newData[,1])) +  stat_qq() + stat_qq_line()

#THE SIMULATED DISTRIBUTION APPRACHES THE NORMAL DISTRIBUTION AS THE SAMPLE SIZE INCREASES

newMeans2 = NULL
for (i in 1 : 1000) newMeans2 = c(newMeans2, mean(rexp(60, lambda)))
newData2 <- data.frame(newMeans2,size=60)

#The distribution of means approaches a normal distribution as the size of the sample
#increases.
newp2 <- ggplot(newData2, aes(newMeans2, fill = 30)) + theme_bw() +
         geom_histogram(aes(y = ..density..), alpha = 0.7, binwidth = 0.30,
                col = "black")  +
         stat_function(mapping = NULL, data = NULL, geom = "path",
                position = "identity",
                fun=dnorm, size = 1, args=list(mean=5, s = sd(newMeans2)))  +
         geom_vline(aes(xintercept=mean(newMeans2),colour="red")) +
         geom_text(aes(x = mean(newMeans2), data = NULL, label="\nmean",y=0.2),
               colour="black",angle=90, size = 3) +
         ggtitle("Smulated distribution verus
normal distribution - 1000 simulations -
sample size 60") +
         theme(plot.title=element_text(face="bold", size=9))


#SAMPLE MEAN VERSUS THEORETICAL MEAN - 1/lambda - 1000simulations n = 60
#mean of means
meanOfMeans <- mean(newMeans2)
cat("Mean of means:" ,meanOfMeans)
```

```
## Mean of means: 4.997003
```

```
#theoretical mean
theoreticalMean <- (1/0.2)
cat("\nTheoretical mean:", theoreticalMean)
```

```
##
## Theoretical mean: 5
```

```
#SAMPLE VARIANCE VERSUS THEORETICAL VARIANCE - 1000 simulations n = 60
sampleVariance2 <- var(newMeans2)
cat("\nSample variance:", sampleVariance2)
```

```
##
## Sample variance: 0.4202573
```

```
theoreticalVariance <- ((1/0.2)/sqrt(40))^2
cat("\nTheoretical variance:", theoreticalVariance)
```

```
##
## Theoretical variance: 0.625
```

```
#Quantile-Quantile plot
newqq <- ggplot(newData, aes(sample = newData[,1])) +  stat_qq() + stat_qq_line()

#SAMPLE MEAN VERSUS THEORETICAL MEAN - 1/lambda - 1000 simulations n = 60
#mean of means
meanOfMeans2 <- mean(newMeans2)
cat("Mean of means:" ,meanOfMeans2)
```

```
## Mean of means: 4.997003
```

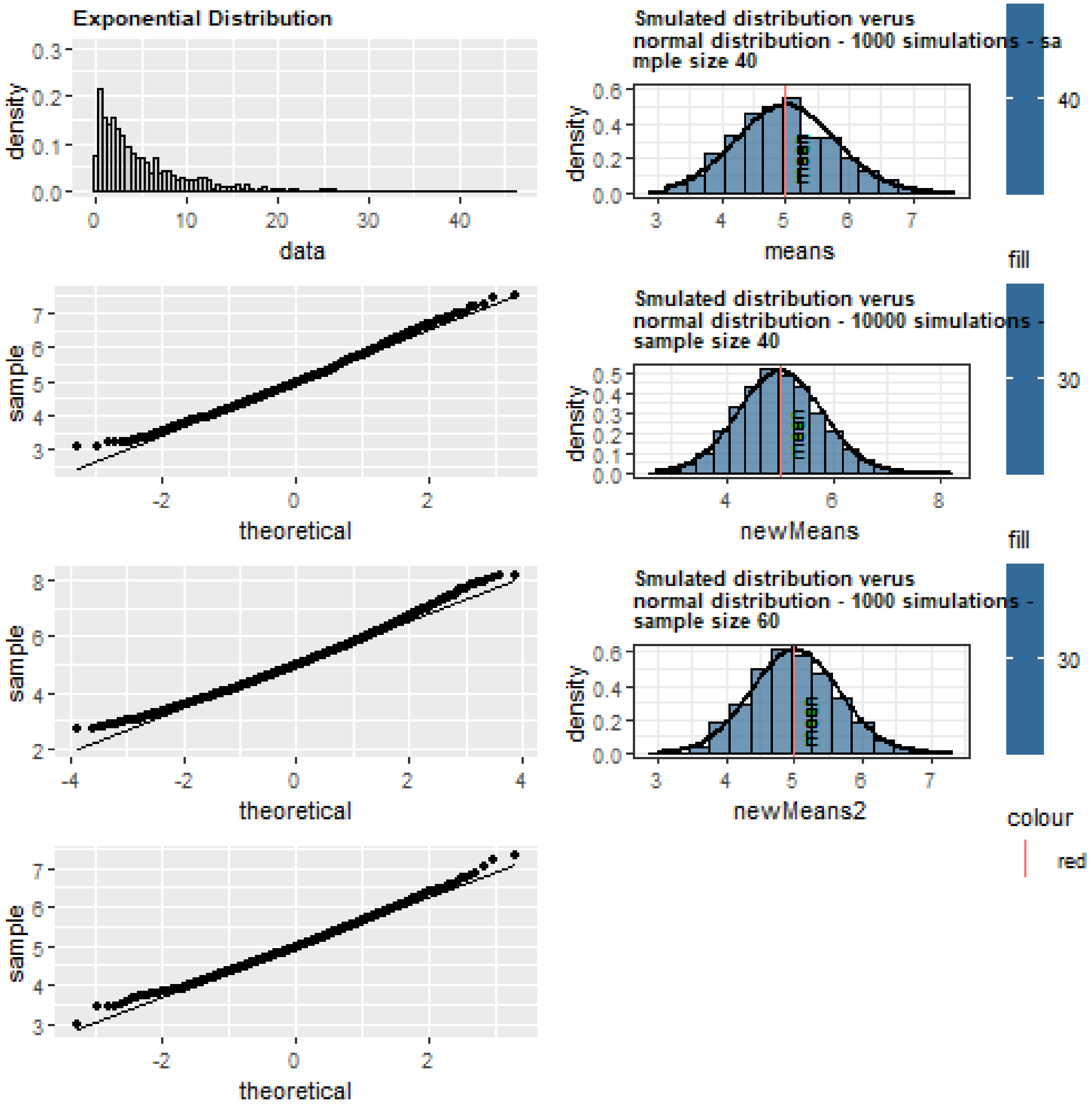
```
#theoretical mean
theoreticalMean <- (1/0.2)
cat("\nTheoretical mean:", theoreticalMean)
```

```
##
## Theoretical mean: 5
```

```
#Quantile-Quantile plot
newqq2 <- ggplot(newData2, aes(sample = newData2[,1])) +  stat_qq() + stat_qq_line()

grid.arrange(ex, p, qq, newp, newqq, newp2, newqq2, nrow = 4)
```

```
#SAMPLE CONFIDENCE INTERVAL VERSUS THEORETICAL CONFIDENCE INTERVAL - 1000 simulation -
#n = 40
sampleConfidenceInterval <- round(mean(means) + c(-1,1)*1.96*sd(means)/sqrt(sample), 3)
```



Exponential Distribution

Smulated distribution verus normal distribution - 1000 simulations - sa mple size 40

Smulated distribution verus normal distribution - 10000 simulations - sample size 40

Smulated distribution verus normal distribution - 1000 simulations - sample size 60

Figure 1: plot of distributions

```r
cat("\nsample confidence interval - 1000 simulations - n = 40: ", sampleConfidenceInterval)
```

```
##
## sample confidence interval - 1000 simulations - n = 40:  4.748 5.232
```

```r
theoreticalConfidenceInterval <- theoreticalMean + c(-1,1)*1.96*theoreticalVariance/sqrt(sample)
cat("\ntheoretical confidence interval: ", theoreticalConfidenceInterval)
```

```
##
## theoretical confidence interval:  4.80631 5.19369
```

```r
#SAMPLE CONFIDENCE INTERVAL VERSUS THEORETICAL CONFIDENCE INTERVAL - 10000 simulation -
#n = 40
sampleConfidenceInterval <- round(mean(newMeans) + c(-1,1)*1.96*sd(newMeans)/sqrt(sample), 3)
cat("\nsample confidence interval 10000 simulations - n = 40: ", sampleConfidenceInterval)
```

```
##
## sample confidence interval 10000 simulations - n = 40:  4.759 5.247
```

```r
theoreticalConfidenceInterval <- theoreticalMean + c(-1,1)*1.96*theoreticalVariance/sqrt(sample)
cat("\ntheoretical confidence interval: ", theoreticalConfidenceInterval)
```

```
##
## theoretical confidence interval:  4.80631 5.19369
```

```r
#SAMPLE CONFIDENCE INTERVAL VERSUS THEORETICAL CONFIDENCE INTERVAL - 1000 simulation -
#n = 60
sampleConfidenceInterval <- round(mean(newMeans2) + c(-1,1)*1.96*sd(newMeans2)/sqrt(sample), 3)
cat("\nsample confidence interval 1000 simulations n = 60: ", sampleConfidenceInterval)
```

```
##
## sample confidence interval 1000 simulations n = 60:  4.796 5.198
```

```r
theoreticalConfidenceInterval <- theoreticalMean + c(-1,1)*1.96*theoreticalVariance/sqrt(60)
cat("\ntheoretical confidence interval: ", theoreticalConfidenceInterval)
```

```
##
## theoretical confidence interval:  4.841853 5.158147
```

```r
#CONCLUSSION
#This project is an illustration of Central Limit Theorem which states that the
#distribution ofaverages of independent and identically distributed random variables
#approaches that of a standard normal distribution as the sample size increases
#where the  mean is approximately equal to the population mean and variance equal
#to sigma squared (the polayion variance) over n. Additonally, as the number of samples
#increases to infinity, the distribution of means approaches the normal distribution.
```