# How do Sample Statistics Compare to Theoretical Statistics?

*Ilan Grapel*

*12/26/2016*

## R Markdown

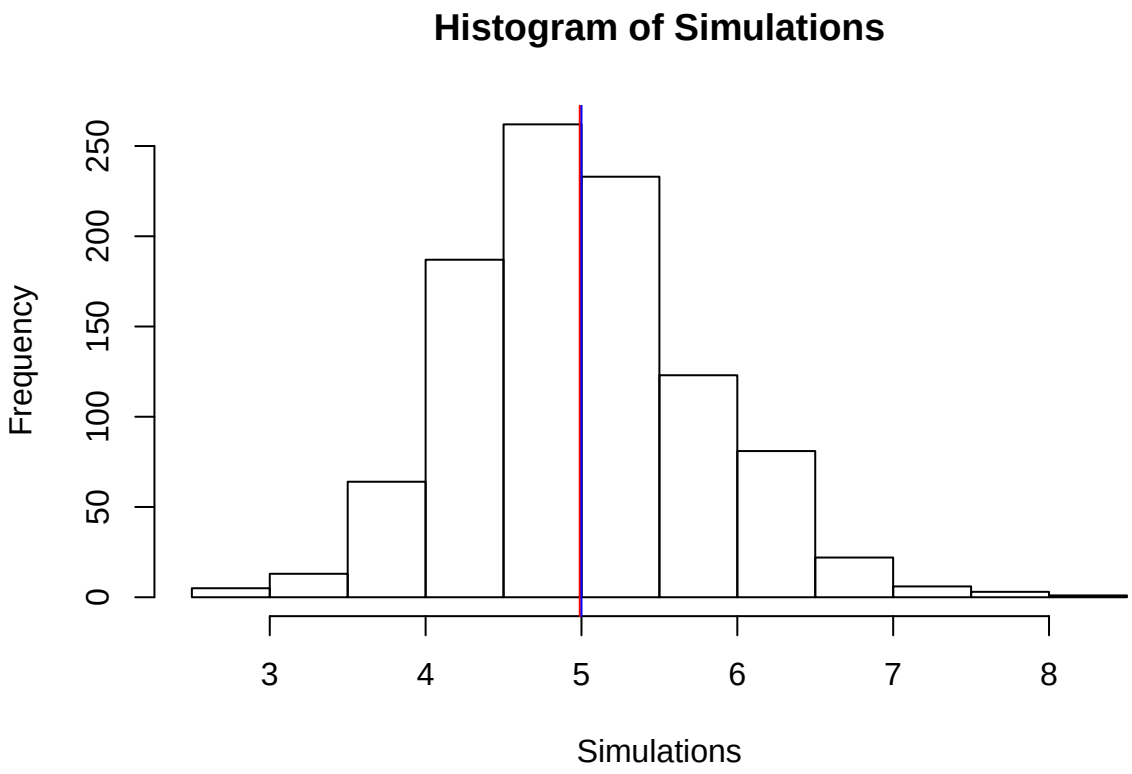
**Overview**: This report is going to examine simulations in the R programming language. Subsequently, the report will compare how the simulation statistics compare to theoretical values.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this

### i. How does the sample mean of an exponential distribution compare to the theoretical mean?

**Methodology**: R will perform 1000 simulations of an exponential distribution with 40 observations and a lamda value of .2. The relevant code for these simulations is:

```
#Ensure consistent values
set.seed(1818)

#Set number of exponentials to 40 and lambda to .2
num <- 40
lambda <- .2

# Run 1000 simulations of the mean value of an exponential function
Simulations=NULL
for(i in 1:1000) Simulations =c(Simulations, mean(rexp(num, lambda)))

#Find the mean of the sample exponential distribution
mean_simulations <- mean(Simulations)

#Create a histogram of the mean of the exponentials
#Red vertical line indicates the mean of the simulations
#Blue vertical line indicats the theoretical mean

hist(Simulations)
abline(v=mean_simulations, col="red")
abline(v=1/lambda, col="blue")
```




As the figure shows, the mean of our samples is a very close to the theoretical value [4.988853 for sample mean versus 5 for the theoretical mean].

**ii. How does the sample variance of the exponential distribution compare to its theoretical variance?**

**Methodology**: The report will compare the distribution's variance to the theoretical variance with the following R code:

```
var_simulations<- (var(Simulations))
theor_var_simulations<- (1/lambda)^2/num
variance_comparison<- matrix(c(var_simulations, theor_var_simulations), nrow=1)
dimnames(variance_comparison)=list(c("Variances"), c("Simulations", "Theoretical"))
variance_comparison
```

```
##           Simulations Theoretical
## Variances   0.6135477       0.625
```

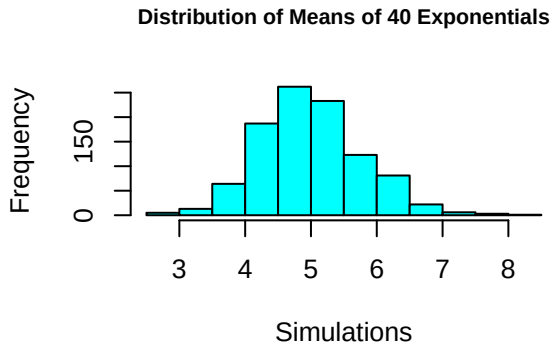
The matrix shows that the empirical variance [.614] is close to the theoretical variance of a mean of 40 exponentials [.625].

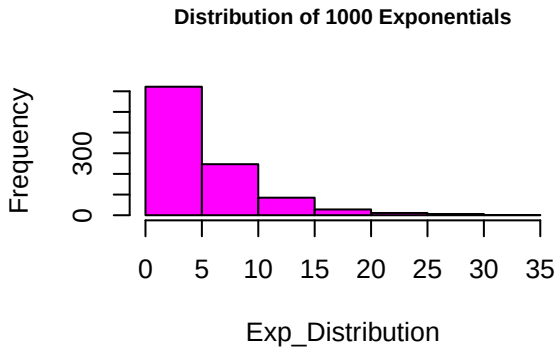
**iii. Is the distribution of the mean of the 40 exponentials approximately normal?**

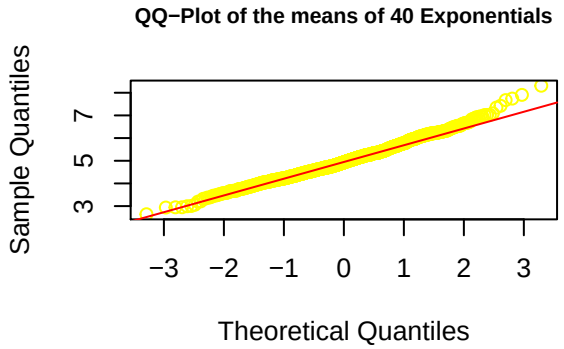
**Methodology**: The report will compare the histograms of the distribution of the mean of 40 exponentials versus the histogram of the distribution of 1000 exponentials.

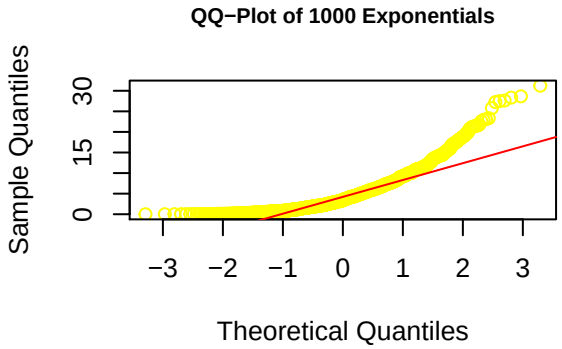
```
#Create a distribution of 1000 exponentials with lambda .2
Exp_Distribution<-rexp(1000,.2)

par(mfrow=c(2,2))
```

```
hist(Simulations, main="Distribution of Means of 40 Exponentials", col=5, cex.main=.8)
hist(Exp_Distribution, main="Distribution of 1000 Exponentials", col=6, cex.main=.8)
qqnorm(Simulations, main="QQ-Plot of the means of 40 Exponentials", col=7, cex.main=.8);
qqline(Simulations, col=2)
qqnorm(Exp_Distribution, main="QQ-Plot of 1000 Exponentials", col=7, cex.main=.8);
qqline(Exp_Distribution, col=2)
```




The histograms and QQ-Plots show that the distribution of the means is approximately normal, in comparison to the exponential distribution, which clearly not.

**Conclusion**: Our observations of 1000 simulations of 40 exponentials show that the distributions of these means adopts a normal distribution. Thus, distribution of 40 exponentials behave as predicted by the Central Limit Theroem.