# Statistical Inference Peer Assessment Part 1 of 2

*Joe Cannon*

*October 19, 2015*

## Comparing Sample Exponential Distribute with Theoretical from the Central Limit Theorem

## Overview

In this project we will investigate the exponential dist in R and compare it with the Central Limit Theorem. The exponential dist can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential dist is 1/lambda and the standard deviation is also 1/lambda.

1. Show the sample mean and compare it to the theoretical mean of the dist.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the dist.
3. Show that the dist is approximately normal.

## Simulations

*Set lambda ($lambda$) = 0.2 for all of the simulations

*Generate 40 exponentials 1000 times

*Calculate the mean, variance, Std. for both simulations and theoretical

*Compare

```
# load neccesary libraries
suppressWarnings(suppressMessages(library(ggplot2)))
```

```
lambda <- .2 # definded setting of lambda
n <- 40  # number of exponentials
nsims <- 1000 # number of tests
set.seed(1961) # For repeatablility.
```

Generate 1000 of 40 exponential distributions, and calculate the mean of each.

```
xdists <- matrix(rexp(n * nsims, lambda), nrow=nsims,ncol=n)
```

```
xdist_mean <- data.frame(mean=apply(xdists, 1, mean))
summary(xdist_mean)
```

```
##       mean
##  Min.   :2.796
##  1st Qu.:4.418
##  Median :4.929
##  Mean   :4.972
##  3rd Qu.:5.467
##  Max.   :8.087
```

## Sample Mean vs. Theoretical Mean

```
tm <- 1/lambda
```

The theroretical mean is $1/\lambda = 1/0.2 = 5$

Calculate the mean of means.

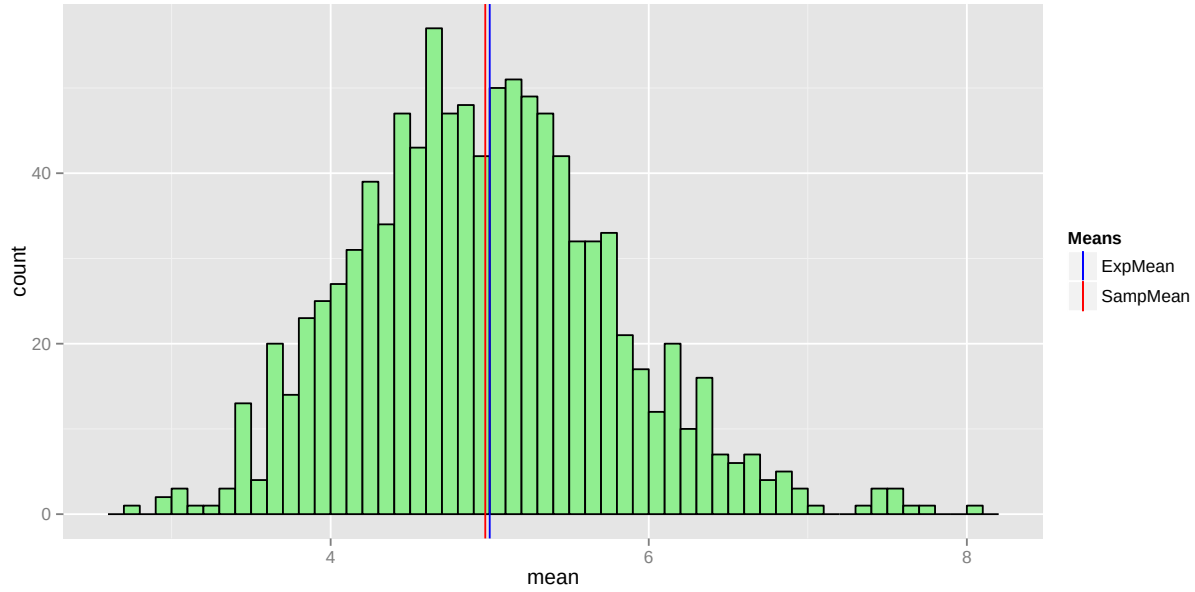
```
xdist_mean_mean <- mean(xdist_mean$mean)
```

Expected mean is 5

Average sample mean is 4.9721775

Observed: They are very close as you can see from the vertical lines on the attached desity plot of sample means

```
ggplot(xdist_mean, aes(x = mean)) +
  geom_histogram(binwidth=0.1,fill="light green", colour="black")+
  geom_vline(aes(xintercept = xdist_mean_mean, colour="SampMean"),show_guide=TRUE) +
  geom_vline(aes(xintercept = tm, colour="ExpMean"), show_guide=TRUE)+
  scale_colour_manual(name="Means", values=c(ExpMean="blue",SampMean="red"))+
   scale_linetype_manual(name="Means", values=c(ExpMean="solid",SampMean="solid"), guide=FALSE)
```




## Sample Variance vs. Theoretical Variance

The expected standard deviation $\sigma$ of a exponential dist of rate $\lambda$ is

```
sd <- 1/lambda/sqrt(n)
```

$\sigma = \frac{1/\lambda}{\sqrt{n}} = \frac{1/0.2}{\sqrt{40}} = 0.7905694$

```
var <- sd^2
```

The expected variance $Var$ of standard deviation $\sigma$ is $Var = \sigma^2 = 0.7905694^2 = 0.625$

```
sdx_mean<- sd(xdist_mean$mean)


varx_mean <- var(xdist_mean$mean)
```

Variance of the average sample mean of 1000 simulations of 40 randomly sampled exponential distribution.

$\bar{Var}_x = 0.6474623$

Corresponding standard deviation.

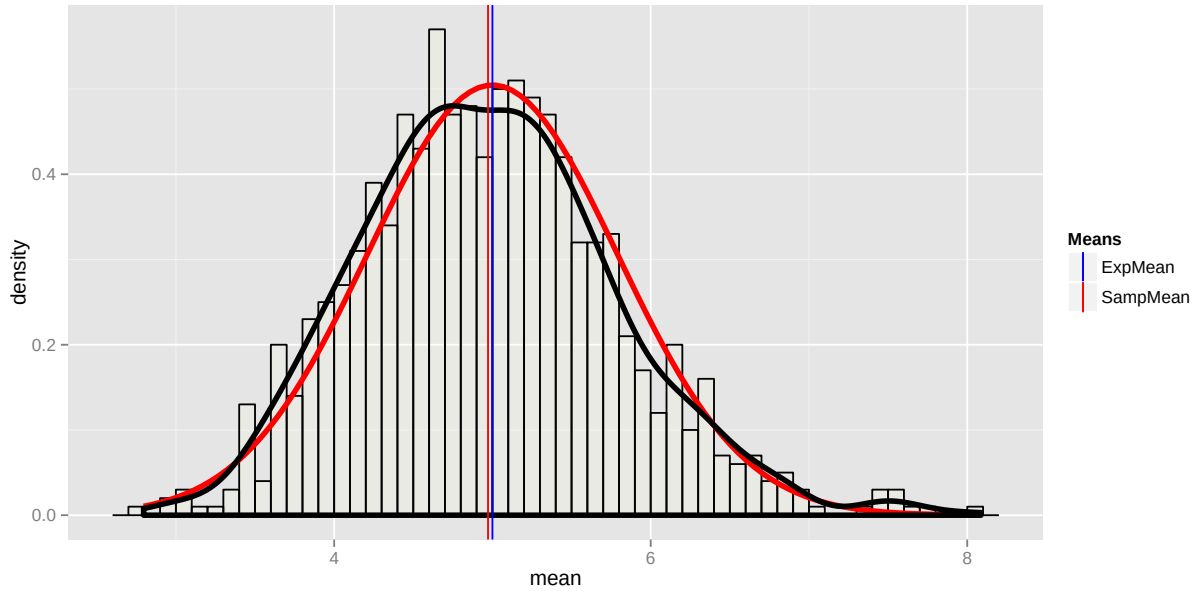
$\bar{\sigma_x} = 0.8046504$

Observation: As you can see, the variances are fairly close, and the standard deviations ar very close.

## Distribution

Comparing the population means & standard deviation with a normal dist of the expected values. Added lines for the calculated and expected means

```
ggplot(xdist_mean, aes(x = mean)) +
  geom_histogram(binwidth=0.1, aes(y=..density..), alpha=0.2,fill="light yellow", colour="black")+
  stat_function(fun = dnorm,
                arg = list(mean = tm , sd = sd),
                colour = "red", size=1.5) +
  geom_density(colour="black", size=1.5) +
  geom_vline(aes(xintercept = xdist_mean_mean, colour="SampMean"),show_guide=TRUE) +
  geom_vline(aes(xintercept = tm, colour="ExpMean"), show_guide=TRUE)+
  scale_colour_manual(name="Means", values=c(ExpMean="blue",SampMean="red"))+
  scale_linetype_manual(name="Means", values=c(ExpMean="solid",SampMean="solid"), guide=FALSE)
```




The red curve shows the Normal Distribution based on the Theroretical mean and standard deviation.

The blue curve shows a smooth density estimate from the sample.

The red vertical line shows the exp distribution mean of means.

The blue vertical line shows the expected mean based on 0.2 lambda

As you can see from the graph, the distribution of means of the sampled exp distribution has a strong fit with the normal distribution of expected mean values, and center on the mean.

As the value lambda increase, the desity will move closer to the center, and lambda becomes smaller we would expect the density to spread out more, but the corralation should remain consistant.