# StatisticalInferenceDMenin

*Diego Menin*

*Friday, October 24, 2014*

## Assignment Part 1

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

### Data Preparation

Fist of All we set the variables and generate 1000 simulations. The function "rexp(n, lambda)" will generate 40 values, and the function "mean" will take their mean. All of that will be done 1000 times using the function replicate.

```
library(reshape2)
library(ggplot2)
library(plyr)


lambda = 0.2
n = 40
sim_number <- 1:1000
rexp_means <- replicate(1000,mean(rexp(n, lambda)))
dfmeans <- data.frame (sim_number, rexp_means)
```

### Questions

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

We were already told that the mean of exponential distribution is 1/lambda, which is:

```
1/lambda
```

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

and is very close to where our distribution is centered (mean of "means of 40 samples from the distributions"), which is:

```
mean(dfmeans$rexp_means)
```

```
## [1] 5.026
```

2. Show how variable it is and compare it to the theoretical variance of the distribution.

According to the Central Limit Theorem, the expected variance can be calculated by (standard deviation / (square root of n)) ^ 2 We are told that the standard deviation of exponential distribution is 1/lambda and n is 40 so the expected variance is:

```r
stddev <- 1/lambda
denominator <- sqrt(n)

expected_variance <- (stddev / denominator) ^ 2
```

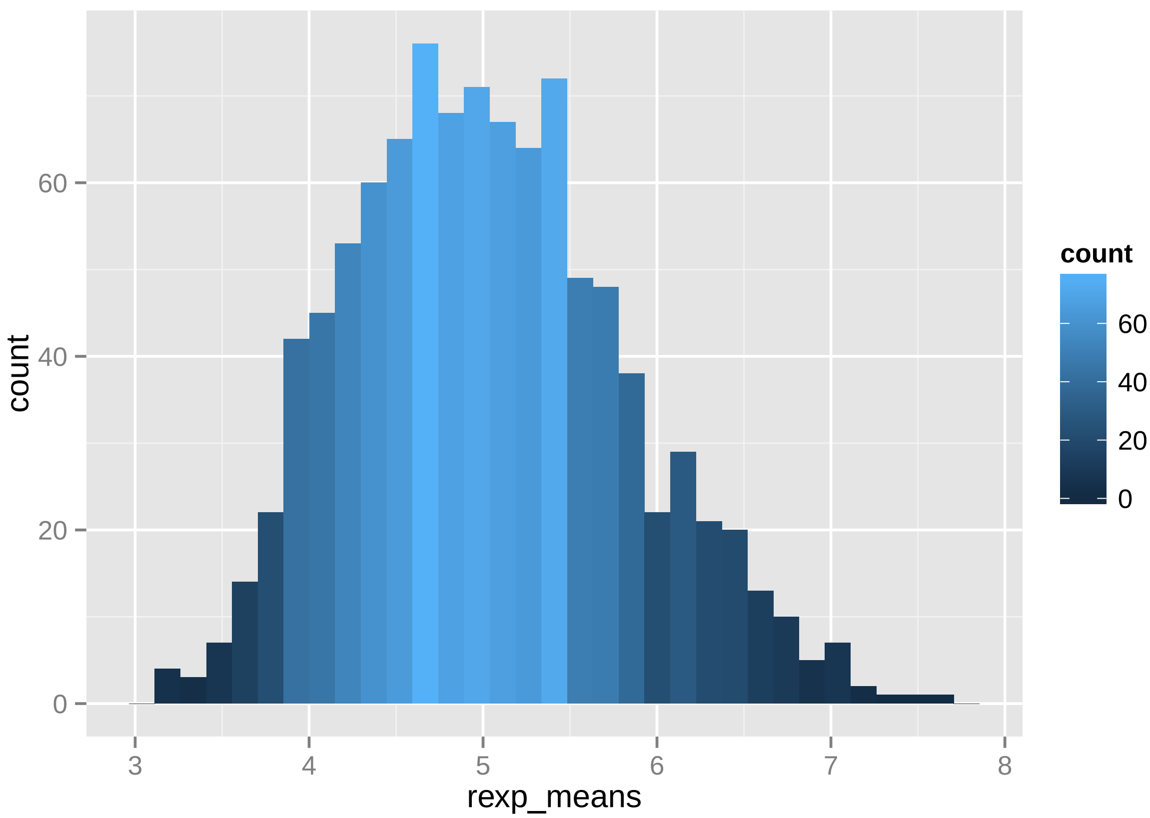
And the actual variance of the distribution is

```r
var(dfmeans$rexp_means)
```

```
## [1] 0.6227
```

3. Show that the distribution is approximately normal.

```r
ggplot(dfmeans, aes(x=rexp_means)) + geom_histogram(aes(fill = ..count..))
```




4. Evaluate the coverage of the confidence interval for 1/lambda: X¯±1.96Sn???.

```r
dfmean <- mean(dfmeans$rexp_means)
standarderror <- sd(dfmeans$rexp_means) /  sqrt(40)

cintervalfrom <- (dfmean) - (1 * 1.96 * standarderror)
cintervalto <- (dfmean) + (1 * 1.96 * standarderror)
```

So the confidence interval is from:

```r
cintervalfrom
```

```
## [1] 4.781
```

to:

```r
cintervalto
```

```
## [1] 5.27
```