# Assignment 1

*Anil Kumar*

*August 20, 2015*

## PART 1

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. Simulation exercises.
2. Basic inferential data analysis. 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.

```
# set lambda = 0.2 for all the simulation
lambda <- 0.2
# set n = 40
n <- 40
```

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal. ## Data Processing

```
setwd("/Users/anilkumar/Desktop/RCOURSE")
library(lattice)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
# initialize the vector to hold the sample distribution
qvector = rep(NA,1000)
# Now store the 1000 simulation in qvector
for(i in 1:1000) {
  temp <- rexp(n,lambda)
  qvector[i] = mean(temp)
}
```

## RESULTS

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

```r
# mean of sample distribution
sample_mean <- mean(qvector)
sample_mean
```

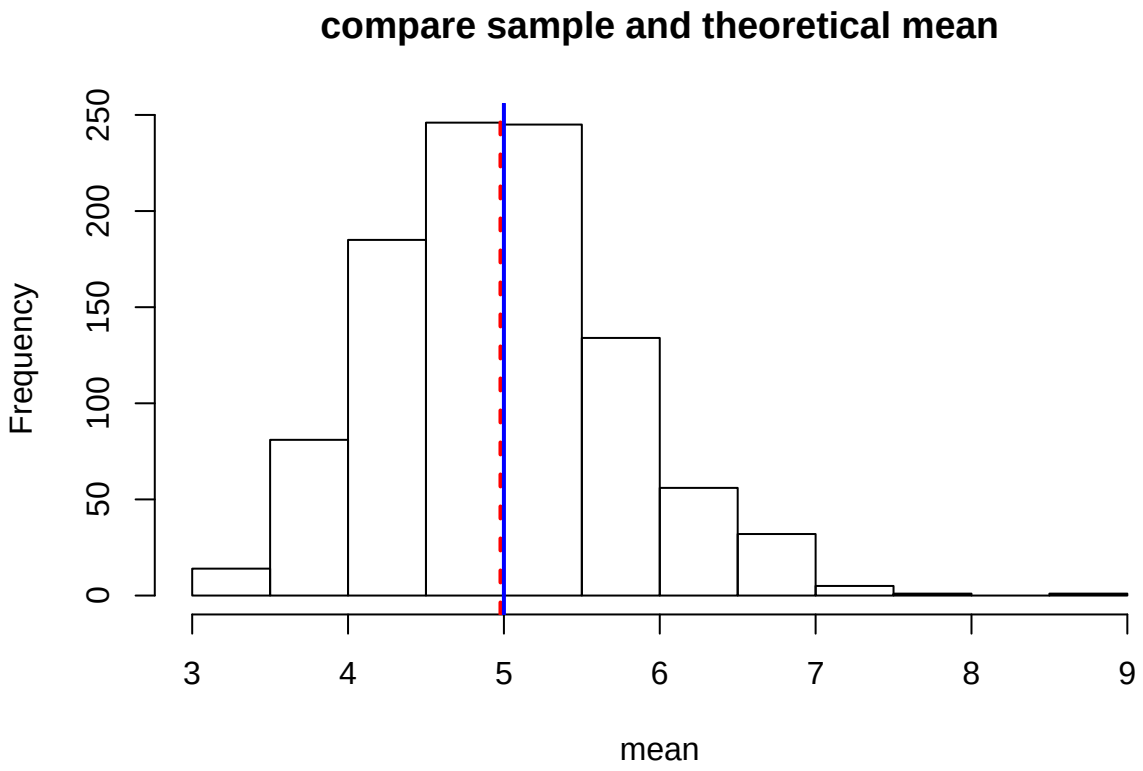
```
## [1] 4.977948
```

```r
# Theoretical center of the distribution
theo_mean <- 1/lambda
theo_mean
```

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

Comparison shows that sample_mean and theoretical center mean, we see mean is very close to theoretical center.

Visualization using plot




CONCLUSION: From above plot, we see that the sample distribution (red dotted line) is very close to our theoretical center of the distribution (blue solid line). Therefore, the center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

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

From the notes, it clear that standard deviaton (SD) is a measure of the evarge deviation around the mean so lets calculate SD of the sample

```r
sample_sd <- sd(qvector)
sample_sd
```

```
## [1] 0.7705296
```

Now calculate the sample variance and compare with theoretical variance

```r
sample_var <- var(qvector)
sample_var
```

```
## [1] 0.5937158
```

Theoretical variance is sigma square divided by n

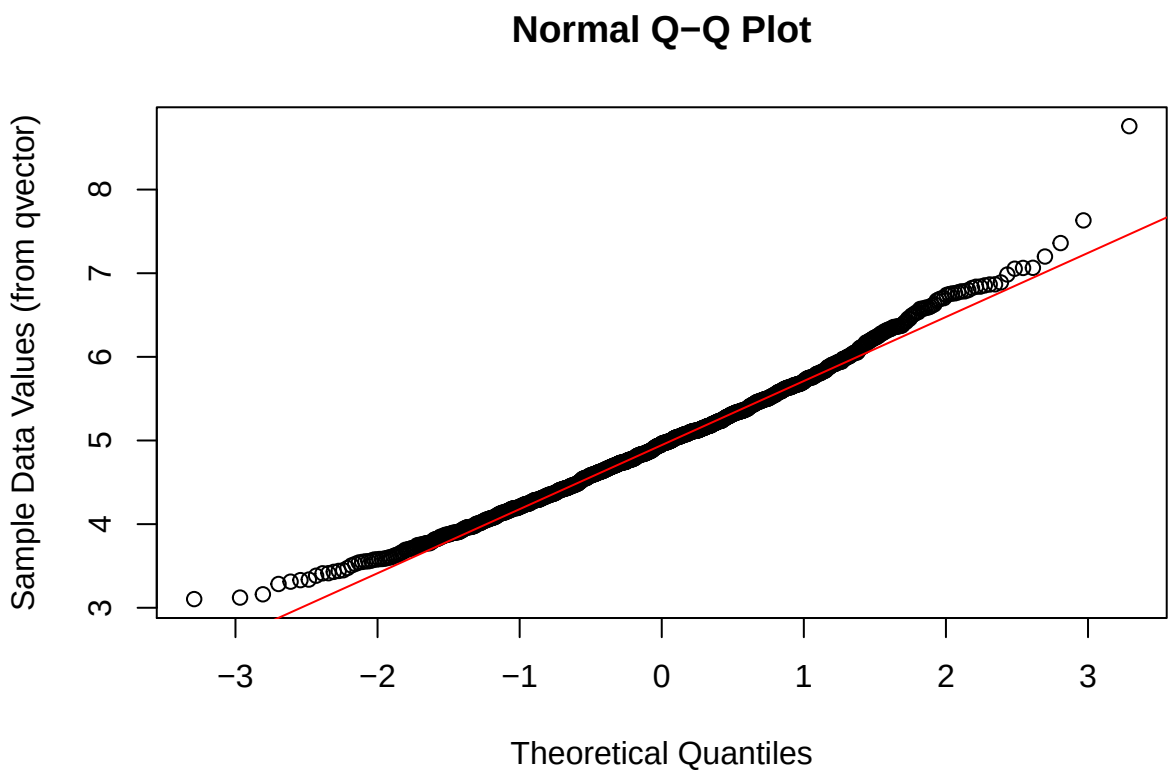
```r
theo_var <- (1/lambda)^2/n
theo_var
```

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

CONCLUSION: The sample variance is close to theoretical variance

**3. Show that the distribution is approximately normal**

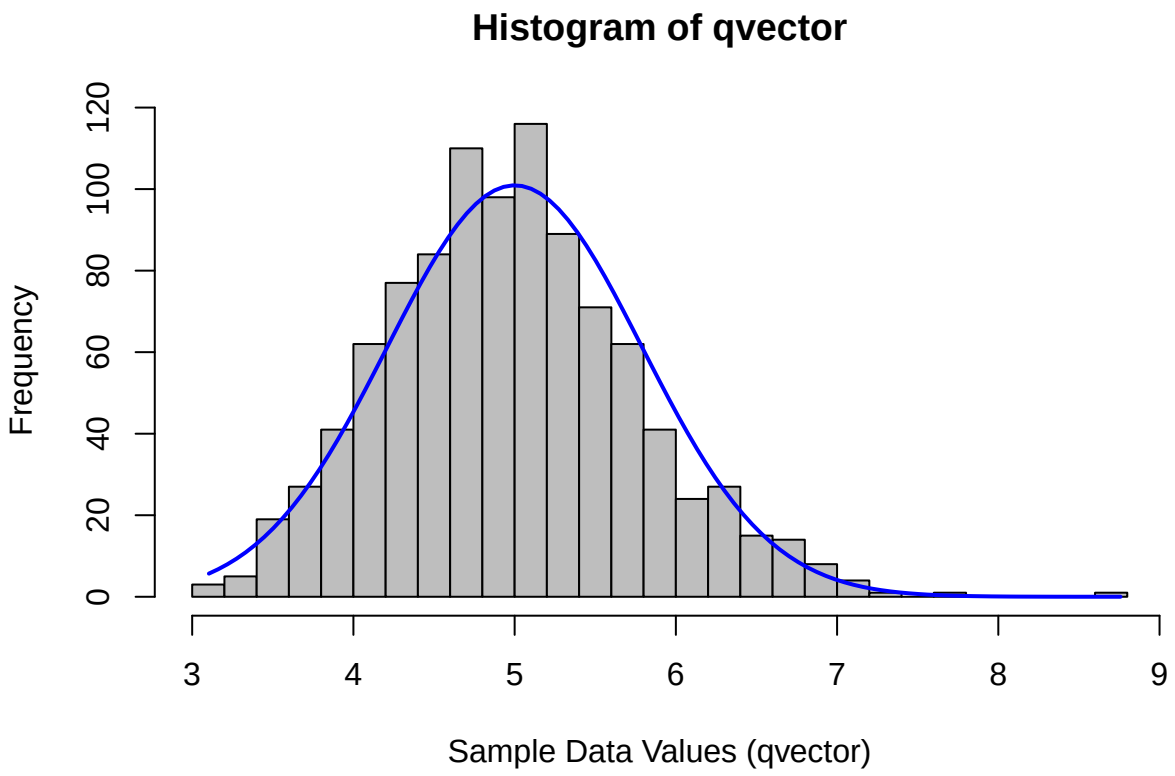
```r
qqnorm(qvector, ylab = "Sample Data Values (from qvector)")
qqline(qvector, col = 2)
```




From the above plot it is clear that distribution is normal as line (red) is quite close fit to the points.

**To show that distribution is normal using histogram**

```
h <- hist(qvector, breaks=n, col= "grey", xlab= "Sample Data Values (qvector)")
xfit <- seq(min(qvector),max(qvector),length=100)
yfit <- dnorm(xfit,mean=theo_mean, sd=sqrt(theo_var))
yfit <- yfit*diff(h$mids[1:2])*length(qvector)
lines(xfit, yfit, col="blue", lwd=2)
```




CONCLUSION We found that distributin is normal.