Statistical Inference Project Part01

Sandeep Anand

March 26, 2017

R Project Heading

The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

Starting with the Simulation Excercise

Exponential Function

- In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem.
- 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 1/lambda.

To-Do's [Step-by-Step]

- Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.
- Show the sample mean and compare it to the theoretical mean of the distribution.
- Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
- Show that the distribution is approximately normal.

```
lambda<-0.2
n=40
B<-1000

samplev<- rep(NA,1000)

## Sample Vector
for (i in 1:1000){
  gen <- rexp(n,lambda)
  samplev[i] <- mean(gen)
}
```

Showing the Sample Mean and Variance and Comparisions between the Sample and the Theoretical distributions

```
library(ggplot2)

#########################
#       Mean            #
#########################
```

```r
sample_mean<-mean(samplev)

given_mean<-1/lambda

vect <- c(sample_mean, given_mean)

sample_mean
```

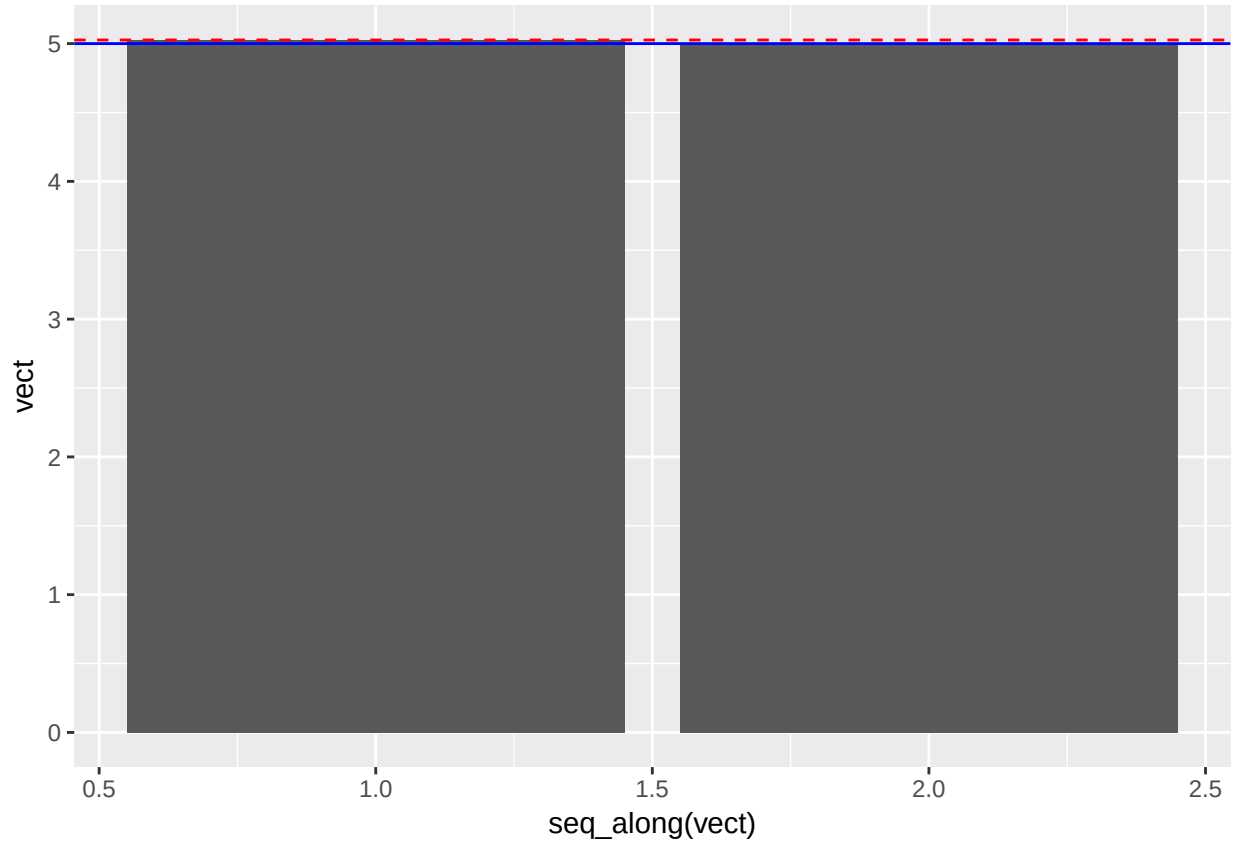
```
## [1] 5.026101
```

```r
given_mean
```

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

```r
g<-ggplot(data.frame(vect), aes(seq_along(vect), vect)) +
  geom_bar(stat = "identity")+
  geom_hline(yintercept = sample_mean, color="red", lty=2) +
  geom_hline(yintercept = given_mean, color="blue")

print(g)
```




```r
#########################
#       Variance        #
#########################

sample_var<-var(samplev)
```

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

vect_var <- c(sample_var, given_var)

sample_var
```

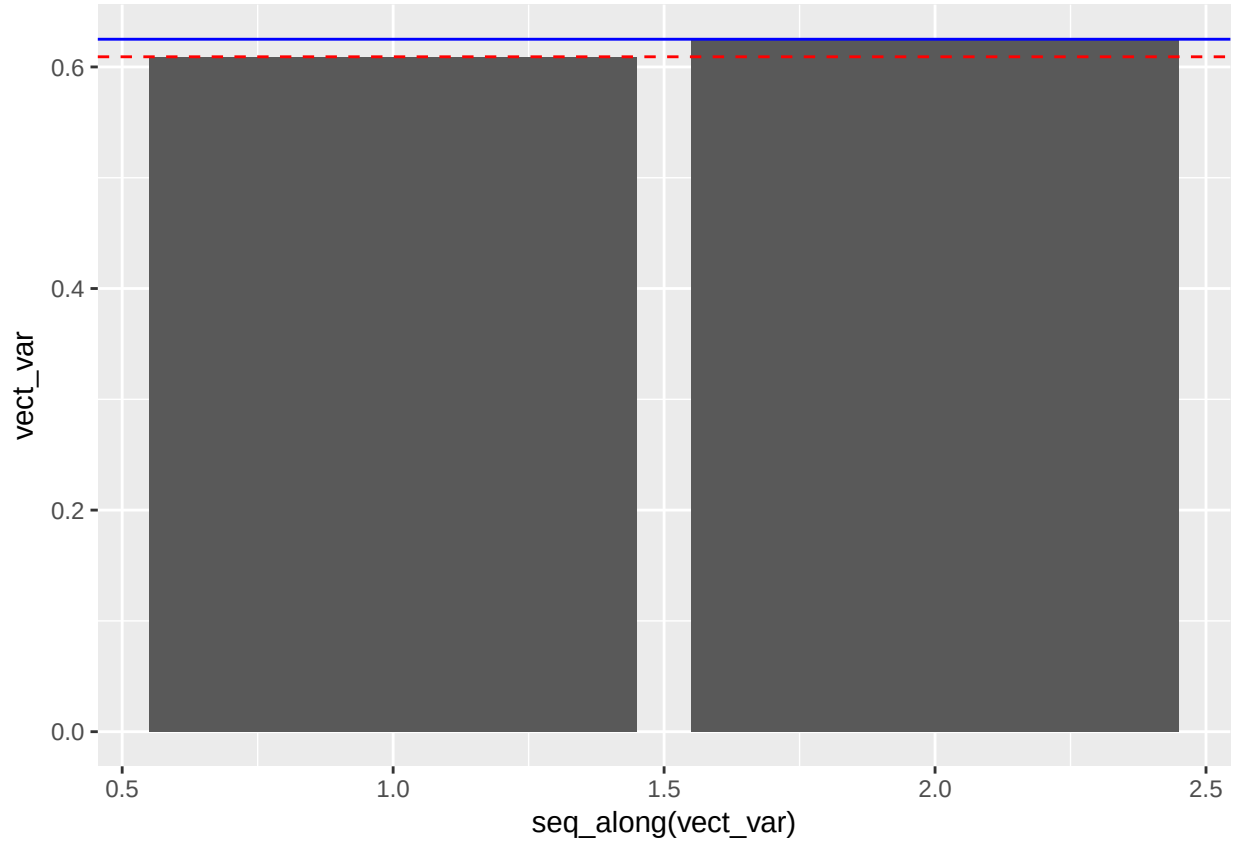
```
## [1] 0.6091626
```

```r
given_var
```

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

```r
g2<-ggplot(data.frame(vect_var), aes(seq_along(vect_var), vect_var)) +
  geom_bar(stat = "identity")+
  geom_hline(yintercept = sample_var, color="red", lty=2) +
  geom_hline(yintercept = given_var, color="blue")

print(g2)
```



Check to show that the distribution is Normal

A random variable is log-normally distributed if its log is a normally distributed random variable and it fits a bell curve

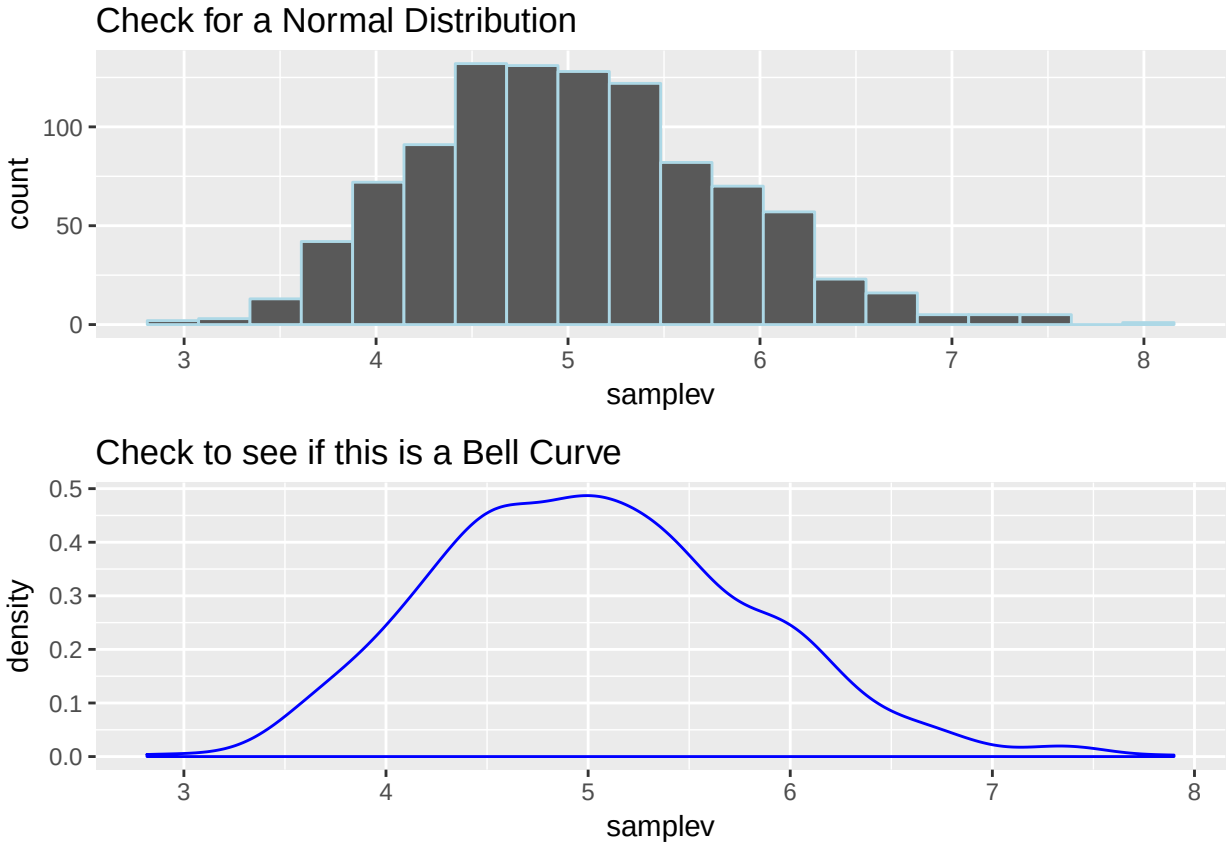
```r
# To See if the disctribution is approximately Normal

g1<-ggplot(data = data.frame(samplev), aes(samplev))+
```

```
    geom_histogram(color="lightblue",bins = 20)+ggtitle("Check for a Normal Distribution")

g3<-ggplot(data=data.frame(samplev),aes(samplev))+
    geom_density(color="blue")+ggtitle("Check to see if this is a Bell Curve")

multiplot(g1, g3)
```



Conclusion

Result : The Distribution is Normal as seen from the Curves above