# Statistical Inference Course Project - Part 1

*Bernardo Najlis*

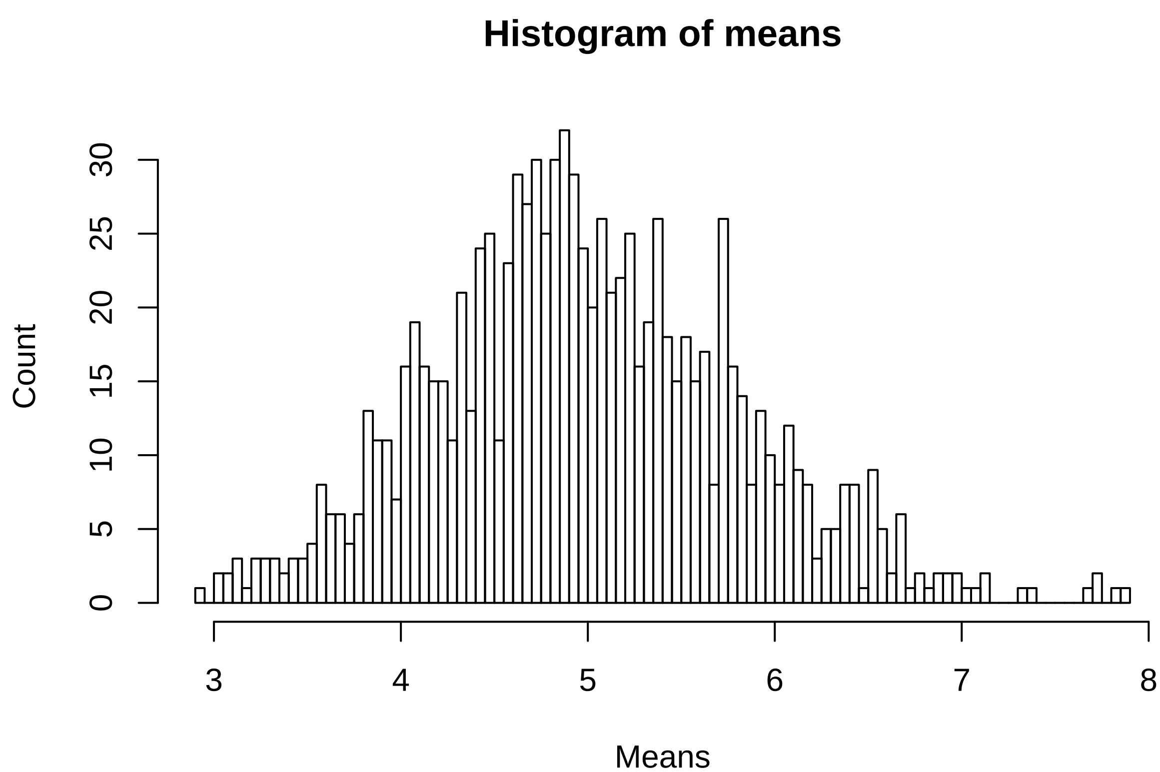
*October 24, 2015*

## Introduction

In this project we 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. We will set lambda = 0.2 for all of the simulations. Also, we will investigate the distribution of averages of 40 exponentials over a thousand simulations.

## Simulations

```r
lambda <- 0.2
n <- 40
numberOfSims <- 1000
set.seed(0) #for reproducibility
expDists <- matrix(data=rexp(n*numberOfSims, lambda), nrow = numberOfSims)
expDistMeans <- data.frame(means=apply(expDists, 1, mean))
```



## Sample and theoretical mean

The expected mean *mu* of an exponential distribution of rate *lambda* is:

*mu* = 1 / *lambda*

```
mu <- 1/lambda
mu
```

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

Then we can calculate the average of sample means for the simulations as

```
meanOfMeans <- mean(expDistMeans$means)
meanOfMeans
```

```
## [1] 4.989678
```

Conclusion: the mean of means converges to the expected mean for a large number of samples (1000).

## Sample and Theoretical variance

The expected standard deviation *sigma* of an exponential distribution rate *lambda* is

*sigma* = (1 / *lambda*) / sqrt(n)

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

```
## [1] 0.7905694
```

The variance *var* of the standard deviation *sigma* is

*var* = *sigma* ^ 2

```
var <- sigma ^ 2
var
```

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

For the simulations, let *var_x* be the variance of the average sample mean of the 1,000 simulations of 40 randomly sampled exponential distributions and *sigma_x* its standard deviation.

```
sigma_x <- sd(expDistMeans$means)
sigma_x
```

```
## [1] 0.8147837
```

```
var_x <- var(expDistMeans$means)
var_x
```

```
## [1] 0.6638725
```

Conclusion: the numbers are close to each other and should get closer if the number of simulations further increases.

## Distribution and approximation to normal

The following chart shows the histogram of samples overlayed with a normal distribution with calculated values and the theoretical normal distribution. The theoretical an calculated distributions have a high degree of overlapping, allowing us to use the calculated one as a good approximation as the Central Limit Theorem states.

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```