

# Statistical Inference Course Project - Part One

---

Vitor Sousa

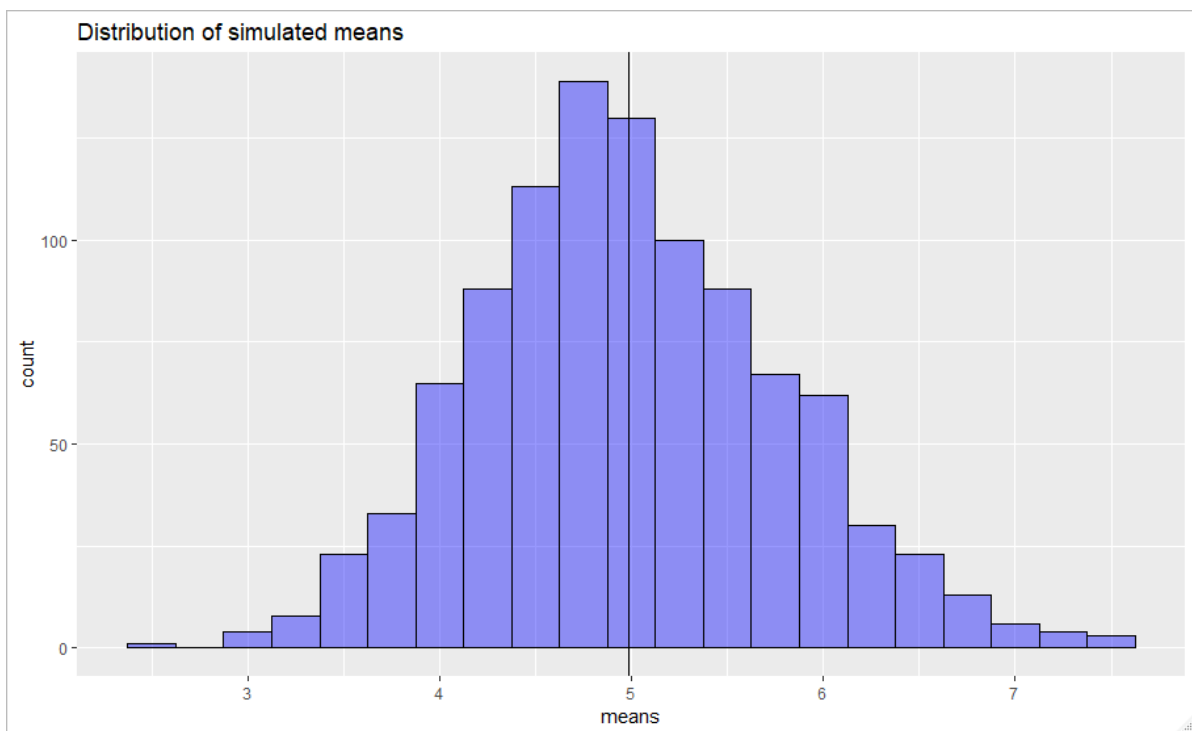
---

The goal of this project is to run a simulation when first we generate 40 variables from an exponential function and repeat this 1000 times. After that, for each simulation, it is computed.

```
#Packages
library(dplyr, warn.conflicts = F)
library(ggplot2)
#Parameters
lambda <- 0.2
n <- 40
num_sim <- 1000
set.seed(119983)
#Create a 1000x40 matrix
sim_distrib <- matrix(data=rexp(n * num_sim, lambda), nrow=num_sim)
```

```
#Mean computation
sim_mns <- data.frame(means=apply(sim_distrib, 1, mean))
sim_mns <- tbl_df(sim_mns)
#Mean of Simulated Means
(mean_sim <- sim_mns %>% summarize(simulated_mean = mean(means)) %>% unlist())
#simulated_mean
#      4.982365

#Plot
sim_mns %>%
  ggplot(aes(x = means)) + geom_histogram(alpha=0.4, binwidth= .25, fill =
"blue", col = "black") +
  geom_vline(xintercept = mean_sim, color="black", size = 0.5) +
  ggtitle("Distribution of simulated means")
```



```
#Variance of sample means
sd_samp <- sim_mns %>% select(means) %>% unlist() %>% sd()
(var_samp <- sd_samp ^ 2)
#[0] 0.628253
#Theoretical variance
(((1/lambda))/sqrt(40))^2
#[1] 0.625
```

The two values, 0.628253 and 0.625, are very close.

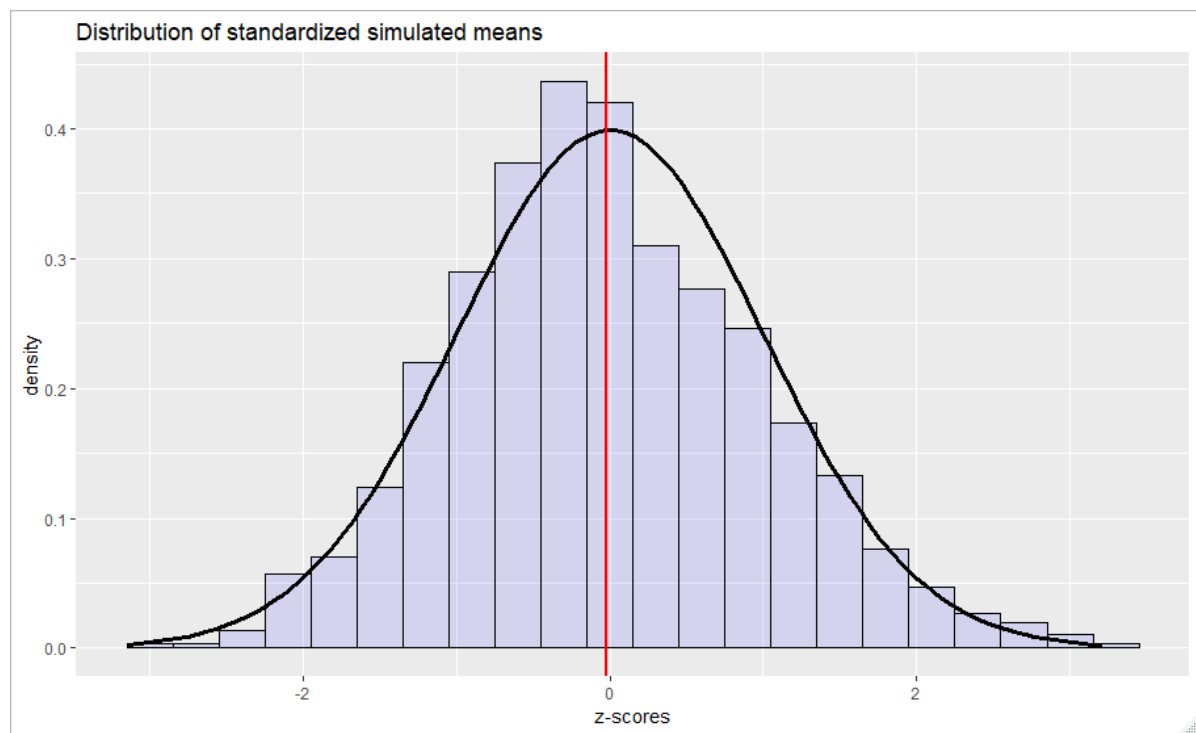
## Normality of the Distribution

From the Central limit theorem we know that the distribution of averages of normalized variables becomes that of a standard normal distribution as the sample size increases.

The CLT defends that the distribution of averages of normalized variables becomes a standard normal distribution with the increase of the sample size.

So I will normalize the sample means. The result, based on CLT, should be a normal distribution centered at 0.

```
(z_mean <- sim_mns %>% mutate(z_score = (means - 1/lambda) / (1/lambda /
sqrt(n))) %>% select(z_score) %>% summarise(z_mean = mean(z_score)) %>%
unlist())
# z_mean
# -0.02230626
#Z-scores
sim_mns %>% mutate(z_score = (means - 1/lambda) / (1/lambda / sqrt(n))) %>%
ggplot(aes(x = z_score)) +
  geom_histogram(alpha=0.1, binwidth = 0.3, fill="blue", color="black", aes(y =
..density..)) +
  stat_function(fun = dnorm, size = 1.3) +
  geom_vline(xintercept = z_mean, color="red", size = 1) +
  ggtitle("Distribution of standardized simulated means") +
  xlab("z-scores")
```



The values and the plot show that the normalized distribution of the sample means is approximately the same as the standard normal distribution. The mean is also very close to 0, which confirms what is deduced from CLT.