Project1

TWW

2024-05-04

Exponential Distribution and Central Limit Theorem

Overview

In this project, we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. We will simulate the distribution of averages of 40 exponentials and examine its properties.

Simulations

We will use the `rexp()` function in R to simulate the exponential distribution. The mean and standard deviation of the exponential distribution are both equal to 1/lambda, where lambda is the rate parameter. We set lambda = 0.2 for all simulations.

To simulate the distribution of averages of 40 exponentials, we will perform 1000 simulations:

```
set.seed(42)
lambda <- 0.2
n <- 40
num_sims <- 1000

sim_means <- replicate(num_sims, mean(rexp(n, lambda)))

#Sample Mean versus Theoretical Mean
#The theoretical mean of the exponential distribution is 1/lambda. Let's compare the sample mean from o

theoretical_mean <- 1/lambda
sample_mean <- mean(sim_means)

paste("Theoretical Mean:", theoretical_mean)
```

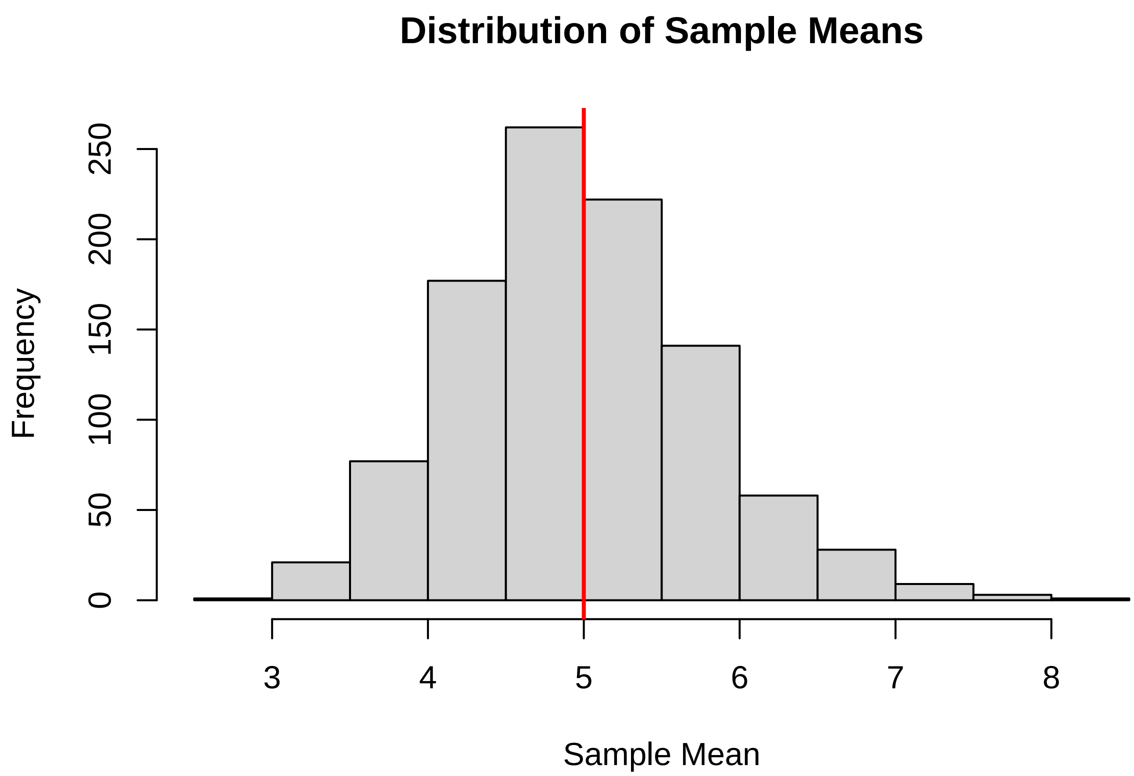
```
## [1] "Theoretical Mean: 5"
```

```
paste("Sample Mean:", sample_mean)
```

```
## [1] "Sample Mean: 4.98650831745453"
```

```
#We can visualize the distribution of sample means using a histogram:

hist(sim_means, main = "Distribution of Sample Means", xlab = "Sample Mean")
abline(v = theoretical_mean, col = "red", lwd = 2)
```




```r
#The red vertical line represents the theoretical mean.
#Sample Variance versus Theoretical Variance
#The theoretical variance of the exponential distribution is (1/lambda)^2. Let's compare the sample var

theoretical_var <- (1/lambda)^2
sample_var <- var(sim_means)

paste("Theoretical Variance:", theoretical_var)
```

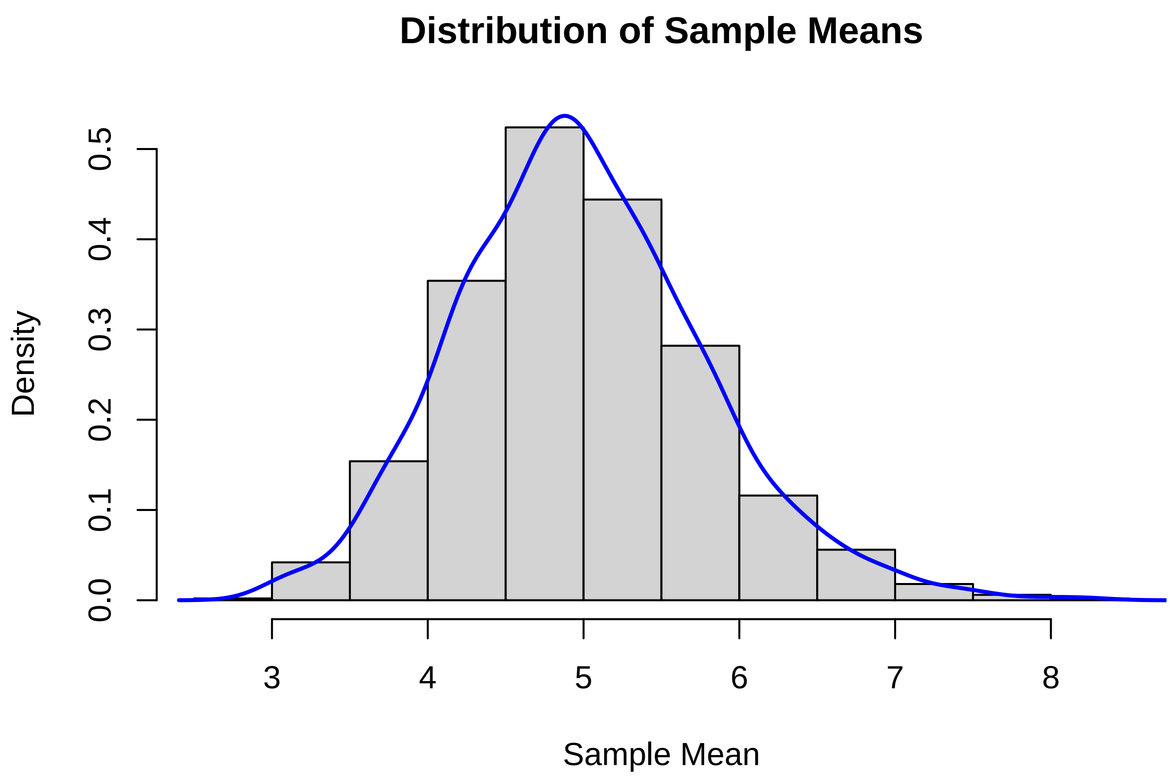
```
## [1] "Theoretical Variance: 25"
```

```r
paste("Sample Variance:", sample_var)
```

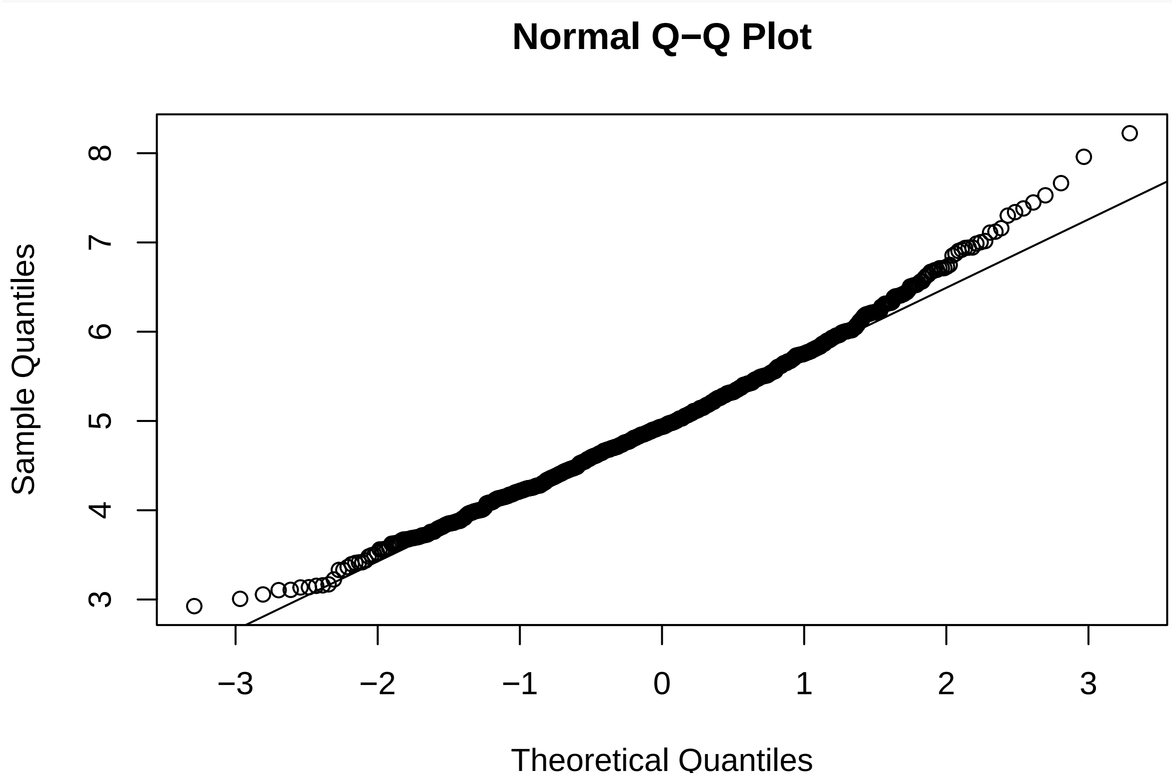
```
## [1] "Sample Variance: 0.634440520668948"
```

```r
#Distribution
#To check if the distribution of sample means is approximately normal, we can create a histogram and a

hist(sim_means, main = "Distribution of Sample Means", xlab = "Sample Mean", prob = TRUE)
lines(density(sim_means), col = "blue", lwd = 2)
```




```r
qqnorm(sim_means)
qqline(sim_means)
```




```r
#If the distribution is approximately normal, the histogram should resemble a bell curve, and the Q-Q p

#Tooth Growth Analysis
```

```r
#Data Summary
#Let's load the ToothGrowth data and perform some basic exploratory data analysis:

data(ToothGrowth)
summary(ToothGrowth)
```

```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000
```

```r
#Confidence Intervals and Hypothesis Tests
#We can use confidence intervals and hypothesis tests to compare tooth growth by supplement type (supp)

# Compare tooth growth by supplement type
t.test(len ~ supp, data = ToothGrowth)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC 
##         20.66333         16.96333
```

```r
# Compare tooth growth by dose
anova(lm(len ~ dose, data = ToothGrowth))
```

```
## Analysis of Variance Table
## 
## Response: len
##           Df Sum Sq Mean Sq F value    Pr(>F)    
## dose       1 2224.3 2224.30  105.06 1.233e-14 ***
## Residuals 58 1227.9   21.17                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Conclusions and Assumptions
#Based on the results of the confidence intervals and hypothesis tests, we can draw conclusions about t
```