Statistical Inference Course Project Part 1

Anas

2024-12-30

Synopsis The project consists of two parts: 1. Simulation Exercise to explore inference 2. Basic inferential analysis using the ToothGrowth data in the R datasets package

Part 1: Simulation Exercise The task is to investigate the exponential distribution (ED) in R and compare it with the Central Limit Theorem. The ED will be simulated in R with rexp(n,lambda) where lambda is the rate parameter. The mean of exponential distribution and the standard deviation are both 1/lambda where lambda = 0.2, and distribution of averages of 40 exponentials and will perform 1000 simulations.

Mean Comparision *Sample Mean vs Theoretical Mean of the Distribution*

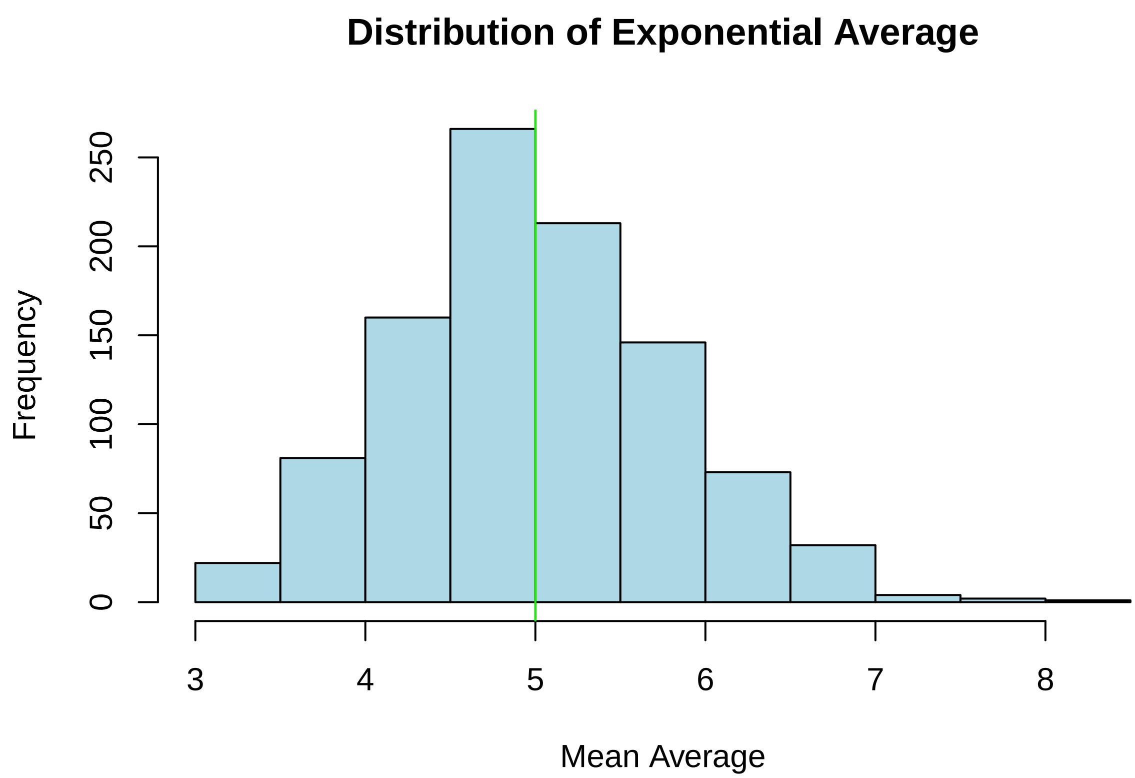
```r
# Sample Mean
sampleMean <- mean(mean_sim_data) # Mean of sample means
print (paste("Sample Mean = ", sampleMean))
```

```
## [1] "Sample Mean =  4.9997019268744"
```

```r
# Theoretical Mean
# the expected mean of the exponential distribution of rate = 1/lambda
theoretical_mean <- (1/lambda)
print (paste("Theoretical Mean = ", theoretical_mean))
```

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

```r
# Histogram shows differences
hist(mean_sim_data, col="light blue", xlab = "Mean Average", main="Distribution of Exponential Average")
abline(v = theoretical_mean, col="brown")
abline(v = sampleMean, col="green")
```



Sample variance vs Theoritical variance

```
# sample deviation & variance

sample_dev <- sd(mean_sim_data)
print (paste("Sample SD = ", sample_dev))
```

```
## [1] "Sample SD =  0.802025077464672"
```

```
sample_variance <- sample_dev^2
print (paste("Sample V = ", sample_variance))
```

```
## [1] "Sample V =  0.643244224882213"
```

```
# theoretical deviation & variance
theoretical_dev <- (1/lambda)/sqrt(n)
print (paste("Theoritical SD = ", theoretical_dev))
```

```
## [1] "Theoritical SD =  0.790569415042095"
```

```
theoretical_variance <- ((1/lambda)*(1/sqrt(n)))^2
print (paste("Theoritical V = ", theoretical_variance))
```

```
## [1] "Theoritical V =  0.625"
```

Distribution

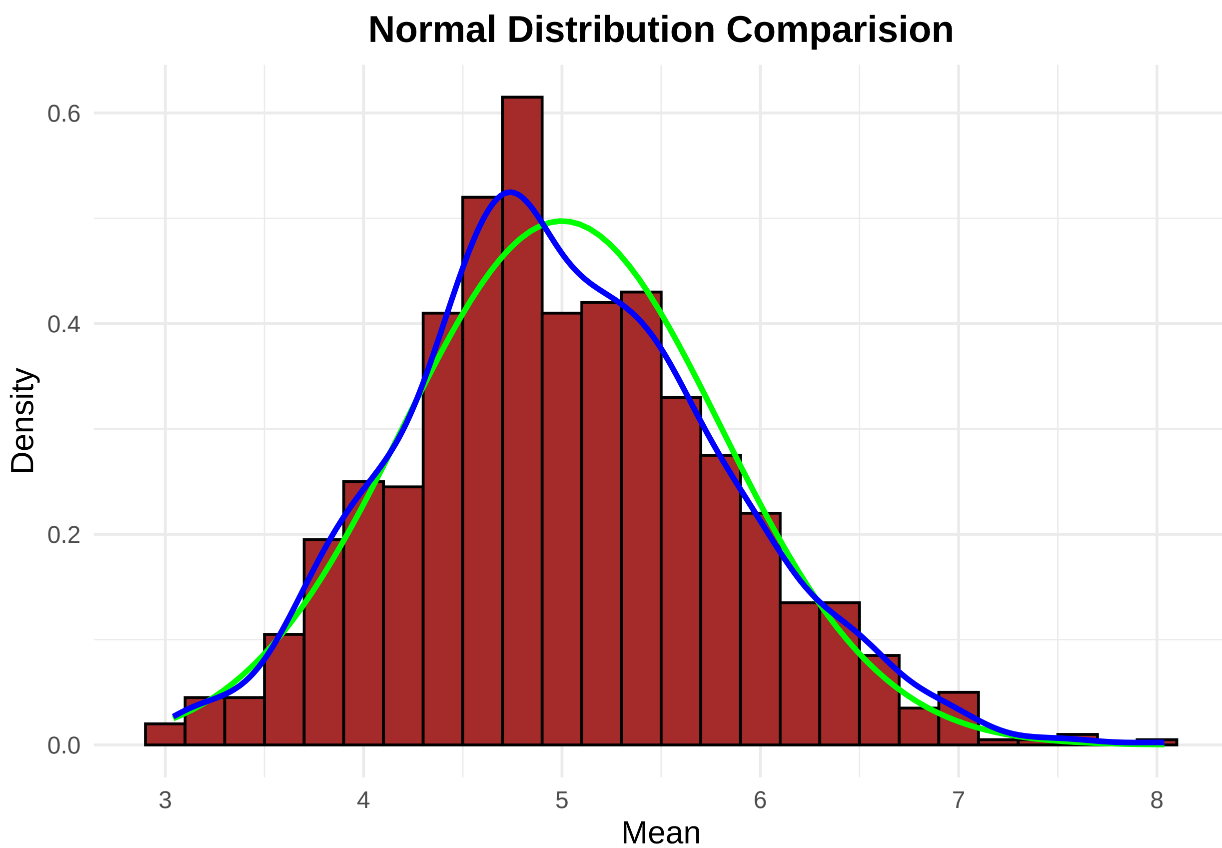
```r
d <- data.frame(mean_sim_data)
t <- data.frame(theoretical_mean)
g <- ggplot(d, aes(x = mean_sim_data)) +

 geom_histogram(binwidth = .2, color="black", fill="brown" , aes(y=..density..))+
           stat_function(fun=dnorm, args=list(mean=theoretical_mean, sd=sd(mean_sim_data)),
                         color="green", size =1) +
           stat_density(geom = "line", color = "blue", size =1)  +
           labs(x="Mean", y= "Density",
                   title="Normal Distribution Comparision")+
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.title = element_text(size = 12)
  )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

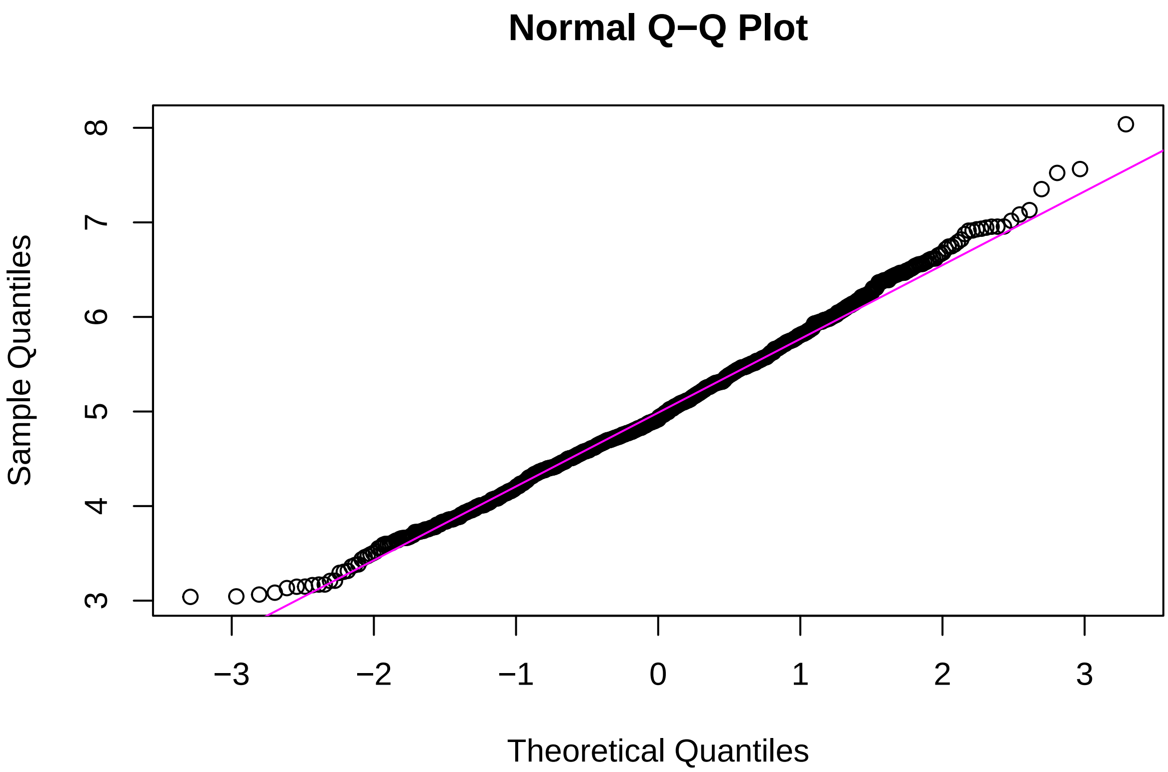
```r
g
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



The above plot indicated that density curve is somehow similar to normal distribution curve. Q-Q Normal Plot also indicates the normal distribution

```
qqnorm(mean_sim_data)
 qqline(mean_sim_data, col = "magenta")
```


Normal Q−Q Plot
Sample Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3
Theoretical Quantiles