StatInference CLT

2024-10-05

R Markdown

This project examines the exponential distribution and its connection to the Central Limit Theorem (CLT) using R.The analysis demonstrates that the distribution of sample means approaches normality, highlighting the differences between individual random exponentials and their averages. This investigation illustrates key statistical principles related to the CLT.

```
# Load necessary library
library(ggplot2)

# Set parameters for the simulation
ECHO = TRUE
set.seed(1337)              # Set seed for reproducibility
lambda = 0.2                # Rate parameter for the exponential distribution
exponentials = 40           # Number of exponentials to average

# Initialize a vector to store simulation means
simMeans = NULL

# Run 1000 simulations to calculate the mean of 40 exponentials each time
for (i in 1:1000) {
  simMeans = c(simMeans, mean(rexp(exponentials, lambda)))  # Generate 40 random exponentials and compu
}

# Sample Mean
mean_simulation = mean(simMeans)  # Calculate the sample mean from simulations
print(mean_simulation)             # Output the sample mean
```

```
## [1] 5.055995
```

```
# Theoretical Mean
theoretical_mean = lambda^-1        # Theoretical mean of the exponential distribution
print(theoretical_mean)              # Output the theoretical mean
```

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

```
# Comparison of Sample Mean and Theoretical Mean
mean_difference = abs(mean(simMeans) - theoretical_mean)  # Calculate the absolute difference
print(mean_difference)             # Output the difference
```

```
## [1] 0.05599526
```

```r
# Sample Variance
sample_variance = var(simMeans)    # Calculate the sample variance from simulation means
print(sample_variance)              # Output the sample variance
```

```
## [1] 0.6543703
```

```r
# Theoretical Variance
theoretical_variance = (lambda * sqrt(exponentials))^-2  # Theoretical variance of the exponential dist
print(theoretical_variance)         # Output the theoretical variance
```

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

```r
# Comparison of Sample Variance and Theoretical Variance
variance_difference = abs(var(simMeans) - theoretical_variance)  # Calculate the absolute difference
print(variance_difference)         # Output the difference
```

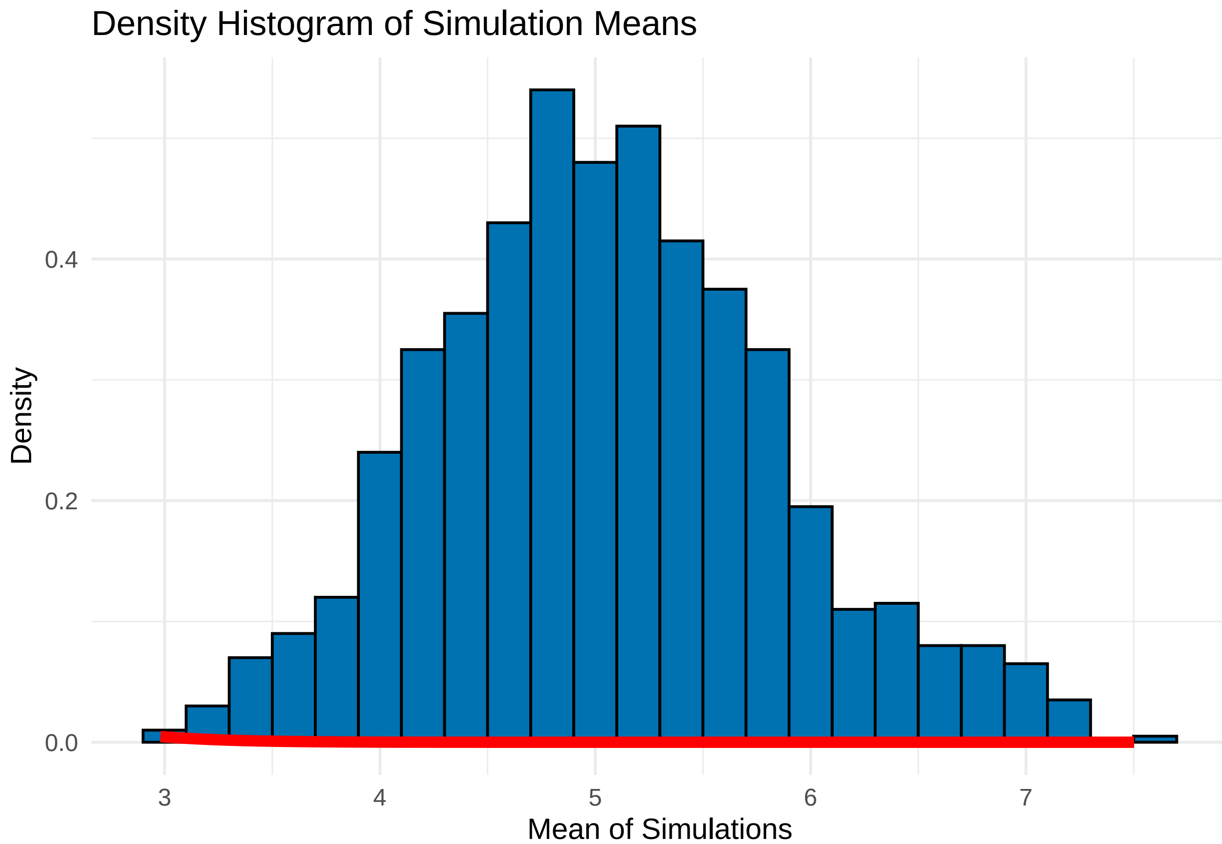
```
## [1] 0.0293703
```

```r
# Distribution
# Create a density histogram of the 1000 simulation means with a normal distribution overlay
ggplot(data.frame(y = simMeans), aes(x = y)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.2, fill = "#0072B2", color = "black") +
  stat_function(fun = dnorm,
              arg = list(mean = theoretical_mean,
                         sd = (lambda * sqrt(exponentials))^-1),
              size = 2, color = "red") +   # Overlay the normal distribution in red
  labs(title = "Density Histogram of Simulation Means",
       x = "Mean of Simulations",
       y = "Density") +
  theme_minimal()                # Apply a minimal theme for better aesthetics
```

```
## 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.
```

```
## Warning in stat_function(fun = dnorm, arg = list(mean = theoretical_mean, :
## Ignoring unknown parameters: `arg`
```

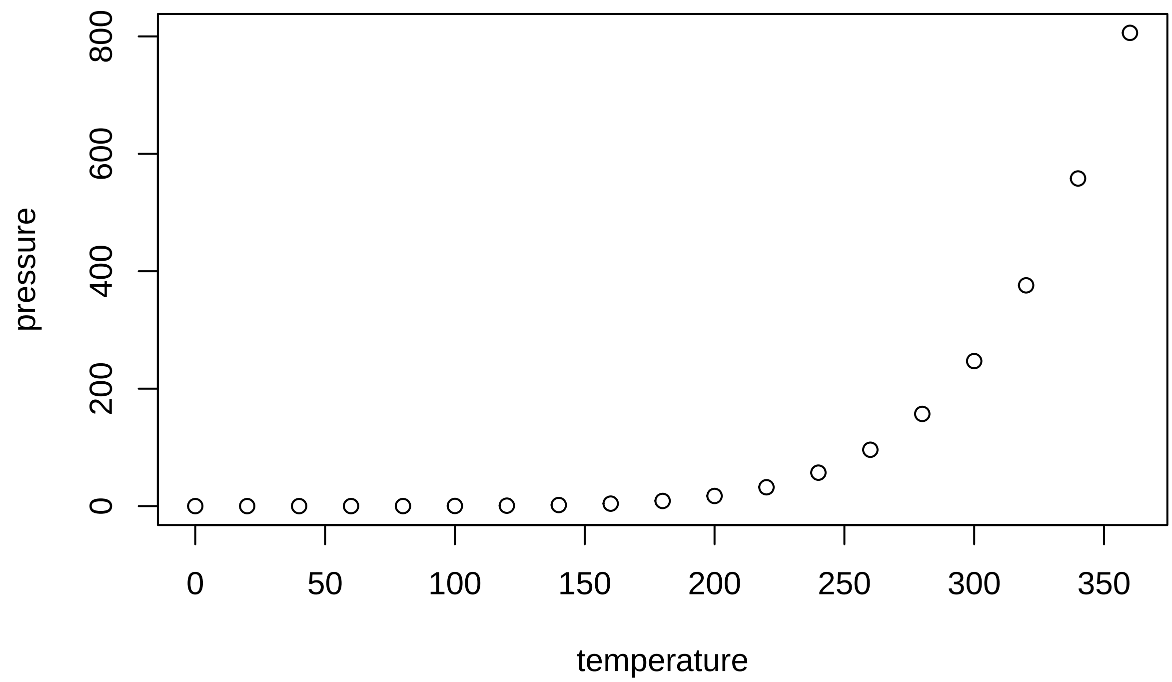
```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



“‘

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.