Statistical Inference Project - Part One

Brandon Robinson

2024-04-13

Overview

The purpose of the project is to investigate two parts relating statistical analysis of data. Part 1 is to perform a simulation on random data and analyze following the Central Limit Theorem. Part 2 uses a dataset from the R dataset library to perform an analysis and derive a conclusion.

Part 1 - Simulation Exercise

Question 1 - Compare the sample and theoretical mean First, the required packages will be loaded in order to run the analysis and create the plots

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Next, the base variables will be set in order to run the simulation

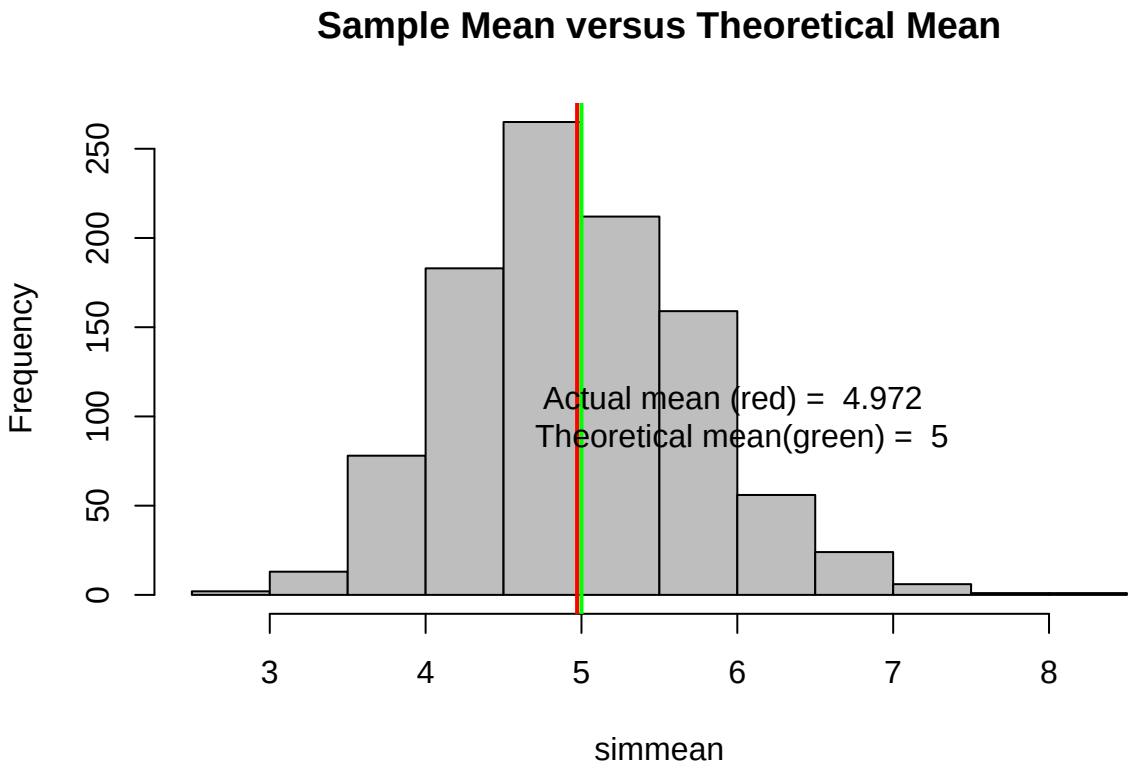
```
#Setting of parameters
set.seed(12345) #provide uniformity of results
lambda <- 0.2
n <- 40
sims <- 1000
```

Then, the simulation will be run in preparation for the plot

```
simrep <- replicate(sims, rexp(n,lambda))
simmean <- apply(simrep, 2, mean)
theormean <- lambda^-1
```

Finally, the plot of the simulation and theoretical mean will be performed

```
hist(simmean, col="gray", main = "Sample Mean versus Theoretical Mean", breaks = 20)
abline(v=mean(simmean), lwd = "2", col = "red")
abline(v=mean(theormean), lwd = "2", col = "green")
text(6,100, paste("Actual mean (red) = ", round(mean(simmean),3), "\n Theoretical mean(green) = ",
                  round(theormean,3)))
```



As can be seen in the plot, the value of the real mean and theoretical mean is small and we can assume the Central Limit Theorem is valid in this scenario.

Question 2: Compare the sampe and theoretical variances The next question asks us to compare the variances and note the difference between the real and theoretical

```
#Sample variance
sampVar <- round(sd(simmean)^2,4)
sampVar
```

```
## [1] 0.5954
```

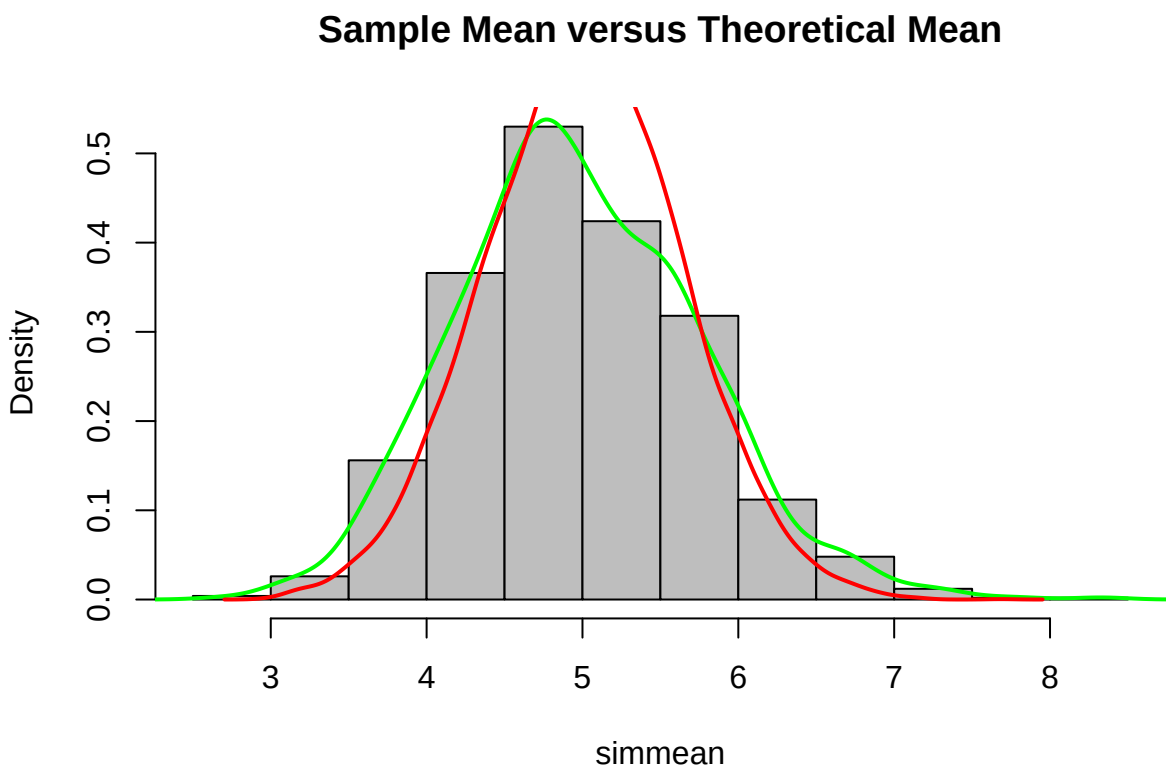
```
#Theoretical variance
theoVar <- round(((1/lambda)/sqrt(n))^2,4)
theoVar
```

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

The difference between the two variance values is quite small.

Question 3: Show the distribution is approximately normal The final step in the process is to show the distribution is approximately normal.

```
hist(simmean, prob=TRUE,col="gray", main = "Sample Mean versus Theoretical Mean", breaks = 20)
lines(density(simmean), lwd = 2, col = "green")
x <- rnorm(10000,mean=5,sd = theoVar)
lines(density(x), lwd = 2, col = "red")
```



Conclusion Based on the simulation data and theoretical results, it can be inferred the random distribution is close to normal