Statistical Inference: Week 4 Project

Author: Alexander Johannes

Date: 10 August 2021

Link to my Statistical Inference repository: My Repo

Overview

This report details the process of carrying out a simulation using the exponential distribution and performing some basic statistical analysis on the `ToothGrowth` data set.

Part I: Simulation Exercise

First, we define some useful variables and load a plotting library.

```
lambda <- .2
samples <- 40
simulations <- 1000
set.seed(7)
library(ggplot2)
```

Next, let's take a sample of 40 exponential distributions of `rate = lambda`, using 1000 simulations, in order to determine the sampling distribution of the mean.

```
simulatedData <- replicate(simulations, rexp(samples, rate = lambda))
```

So, we have a 40 x 1000 matrix containing our simulated data. Let's compute the mean of the sampling distribution and its variability:

```
meansData <- apply(simulatedData, 2, mean)
sampleMean <- mean(meansData)
sampleMean
```

```
## [1] 4.983294
```

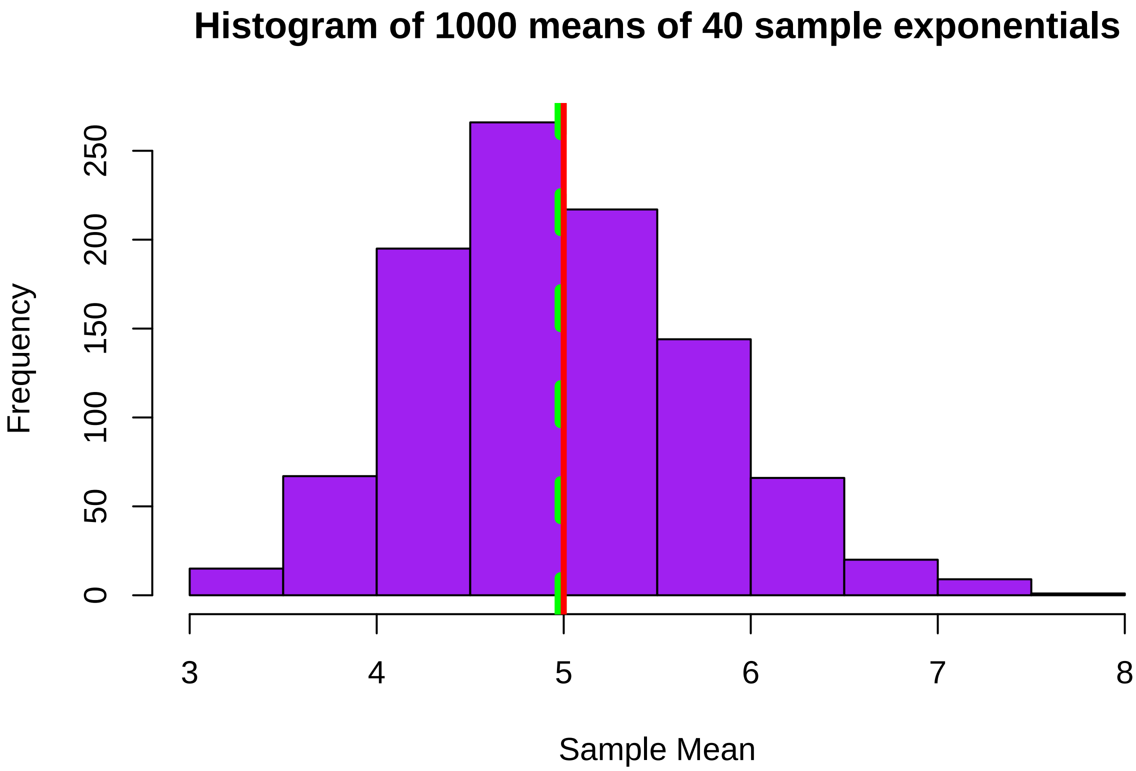
```
theoreticalMean <- 1/lambda

sampleVar <- var(meansData)
sampleVar
```

```
## [1] 0.5792547
```

```
theoreticalVar <- (1/lambda)^2 / samples
```

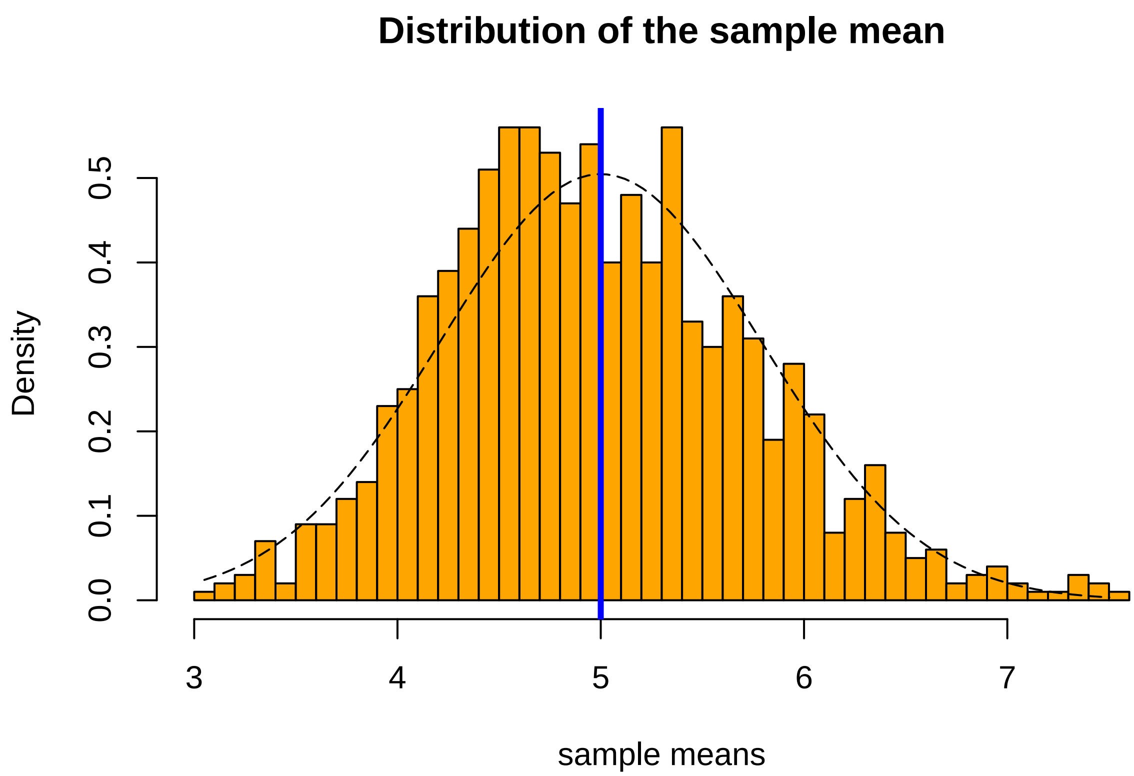
Visualizing our results yield:



Clearly, the sample mean (4.98) and theoretical mean (5) from the simulated exponential distributions are approximately equal.

The sample variance is .58, with the theoretical variance being .625. We expect some variability in our random sample and these values are still fairly similar.

Finally, let's visualize the sampling distribution.



Due to the Central Limit theorem, we see that the distribution of the sample mean is approximately normal. The bell-shaped curve superimposed on the histogram illustrates how the distribution roughly resembles the density function of a normal random variable, centered around the mean (5).

Part II: Basic Inferential Data Analysis

Step 1: Loading the data and performing some basic analysis:

```
data(ToothGrowth)
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

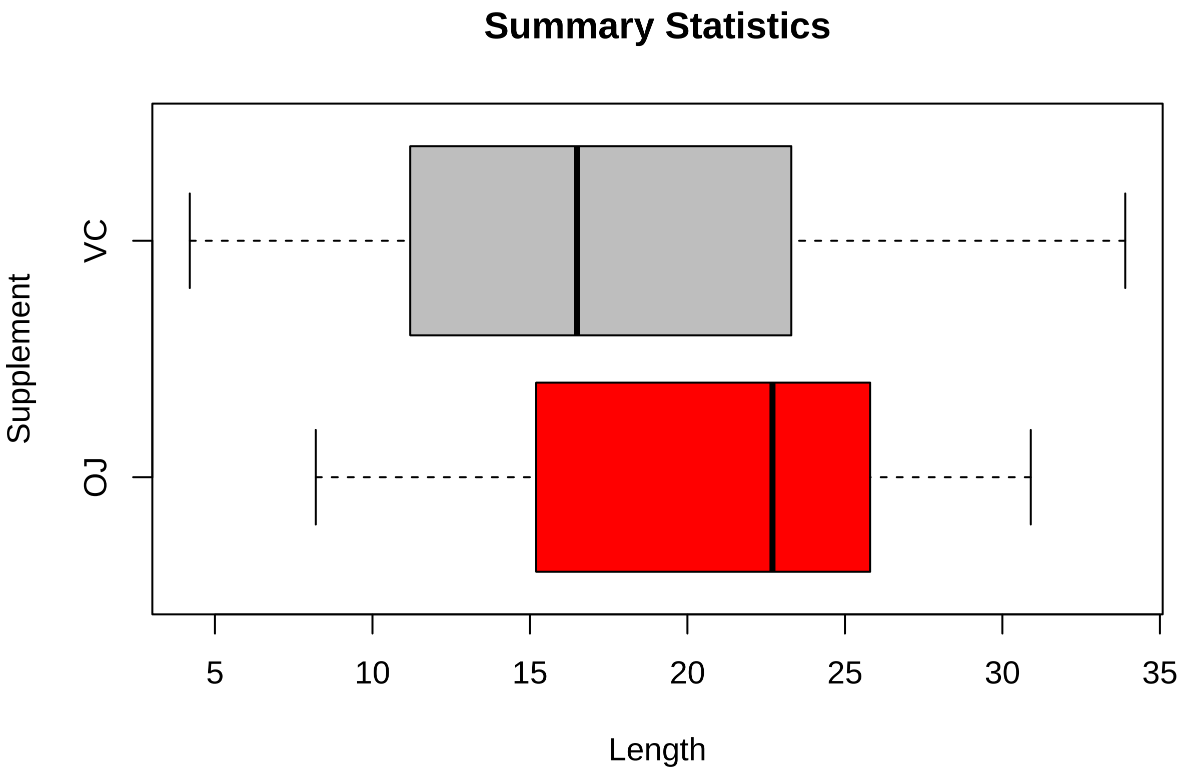
```
tapply(ToothGrowth$len, ToothGrowth$supp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

Step 2: A basic summary of the data, with a visualization:

```
##       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
```



Step 3: Hypothesis tests (t-test) to compare tooth growth by supp and dose

```
##                p.value  Conf.Low Conf.High
## Dosage 0.5 0.006358607  1.719057  8.780943
## Dosage 1   0.001038376  2.802148  9.057852
## Dosage 2   0.963851589 -3.798070  3.638070
```

Step 4: Conclusions and Assumptions

We reject the null hypothesis for a dosage of .5 and 1, meaning there is a difference in tooth growth by delivery method (at the .05 significance level).

We have performed our t-test under the assumption of unequal variances.