Statistical Inference Course Project

Mehdi BENYAHIA

10/03/2020

Overview

Part 1: Simulations

The following was created for part 1 of Courera's Statistical Inference course project. Data was be simulated from the exponential distribution and was used to investigate the Central Limit Theorem. It was found that taking averages of sets of samples of the exponential distribution produced an approximately Normal distribution, while the distribution without means was not approximately normal.

Simulation

Simulation data is created from the exponential distribution.

```
set.seed(150)
lambda <- 0.2
n <- 40
number_simulations <- 1000

simulated_sample <- replicate(number_simulations, rexp(n, lambda))
means_exp <- apply(simulated_sample, 2, mean)
```

Sample Mean vs Theoretical Mean

```
sample_mean <- mean(means_exp)
theo_mean <- 1 / lambda
diff_means <- abs(theo_mean - sample_mean)
```

The diffrence between the Sample Mean and Theoretical Mean is : 0.002 The sample mean and the theoretical mean are very close.

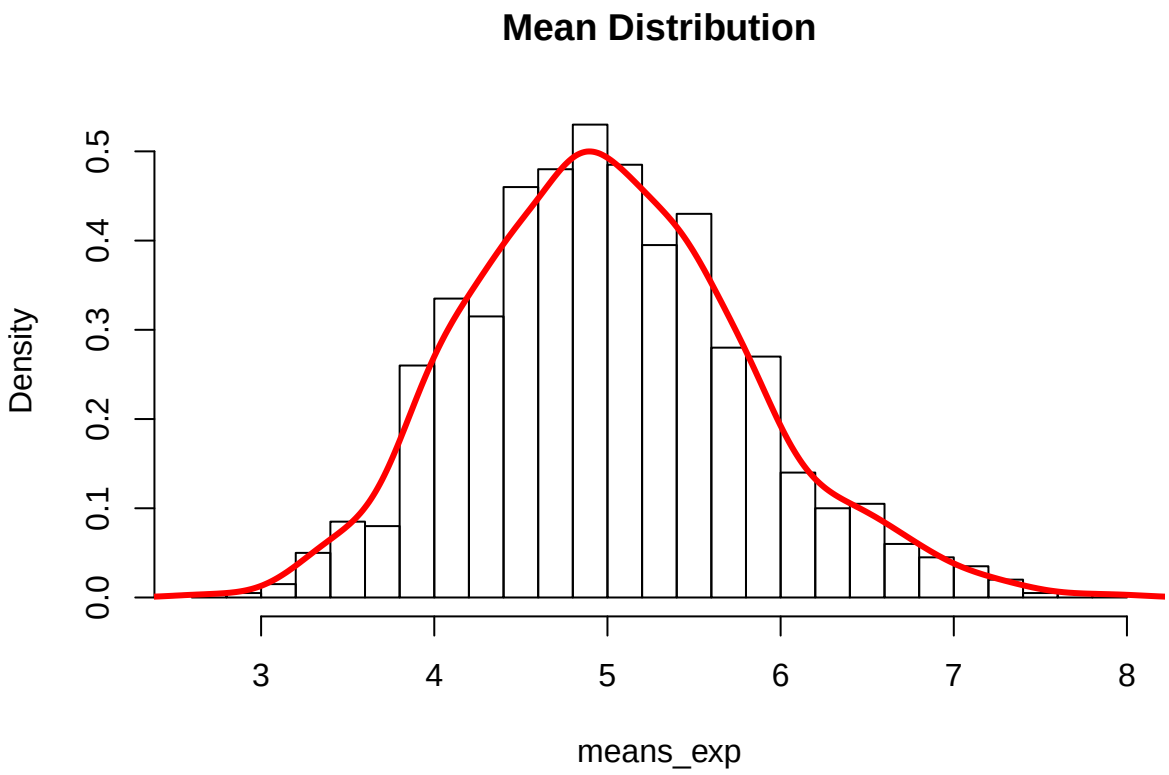
Sample Variance vs theoretical variance

```
sample_var <- var(means_exp)
theo_var  <- (1 / lambda)^2 / (n)
diff_var <- abs(sample_var - theo_var)
```

The diffrence between the Sample Variance and Theoretical Variance is : 0.025 Both variance values are very close to each other.

Distribution

You can also embed plots, for example:



Part 2: Basic Inferential Data Analysis on ToothGrowth Dataset

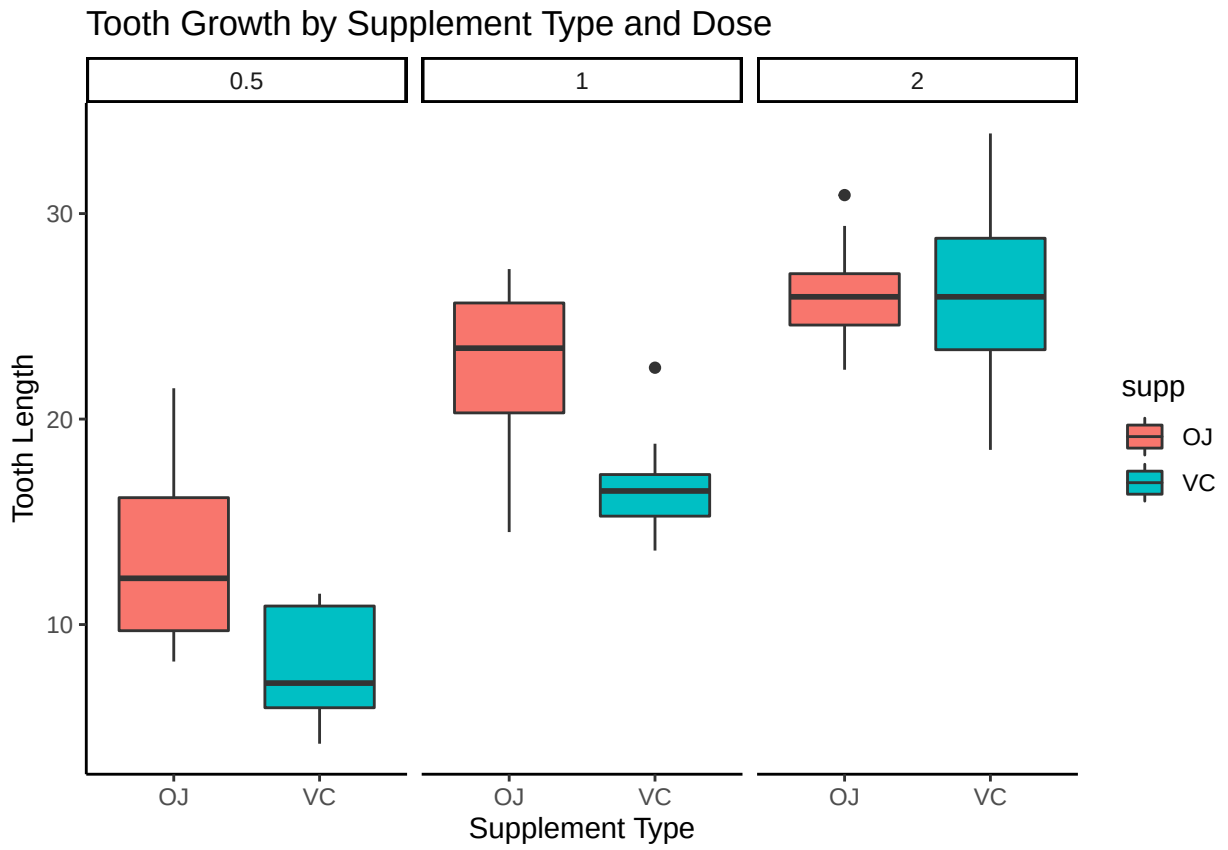
The following was created for part 2 of Courera's Statistical Inference course project [1]. The data from "The Effect of Vitamin C on Tooth Growth in Guinea Pigs" was explored and found to show that increasing dosages of either orange juice or ascorbic acid resulted in longer teeth in the studied guinea pigs. Further, that at all but the highest dosages, orange juice was more effective than ascorbic acid.

Data Summary

```
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
```

Visualize the dataset



Computing Confidence Intervals

```
##   supp dose  mean       sd  n       min      max
## 1   OJ  0.5 13.23 4.459709 10 11.064300 15.39570
## 2   OJ    1 22.70 3.910953 10 20.502166 24.89783
## 3   OJ    2 26.06 2.655058 10 23.730220 28.38978
## 4   VC  0.5  7.98 2.746634 10  5.664568 10.29543
## 5   VC    1 16.77 2.515309 10 14.415961 19.12404
## 6   VC    2 26.14 4.797731 10 23.990112 28.28989
```

```
g <- ggplot(summaries)
g <- g + labs(title = "Comparing length (confidence intervals) by dose", x = "supplement", y = "tooth le
g <- g + geom_crossbar(
  aes(ymin = min, ymax = max, x = supp, y = min, fill = dose),
  fatten = 0
)
g <- g + facet_wrap(~dose)
g
```

Results

Based on the summary above we can say (with 95% of sure that):

- lengths when using OJ (.5 dose) bigger than VC (.5 dose)
- lengths when using OJ (1.0 dose) bigger than VC (1.0 dose)
- lengths when using 2.0 of both OJ and VC are similar.