Statistical Inference

Data Analyst
October 6, 2018

Part 2: Basic Inference

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should Show the sample mean and compare it to the theoretical mean of the distribution. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. Show that the distribution is approximately normal.

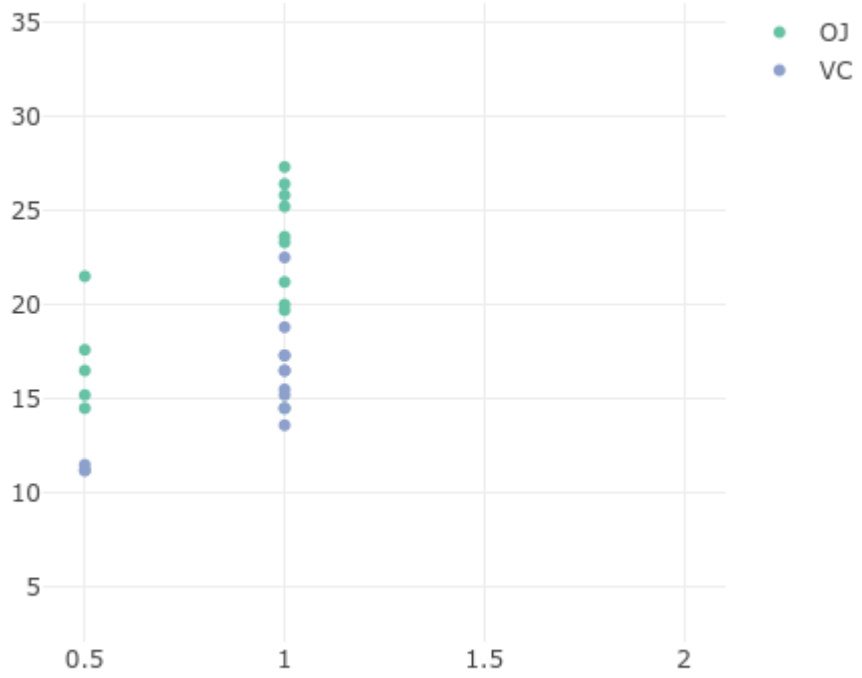
Data Analysis

```
 options(allow_html_dependencies=FALSE)

setwd("/home/asma/Documents/Coursera-DataSciences/Data/course6")
library(datasets)
data(ToothGrowth)
tooth_data<-ToothGrowth
summary(tooth_data)

##       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

head(tooth_data)

##    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(tooth_data)

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

Plot the data and see how it looks like Effect of Dosage and Supplement Type

```
library(plotly)
plot_ly(tooth_data, x= tooth_data$dose, y= tooth_data$len, color =
tooth_data$supp, mode = "markers")
```



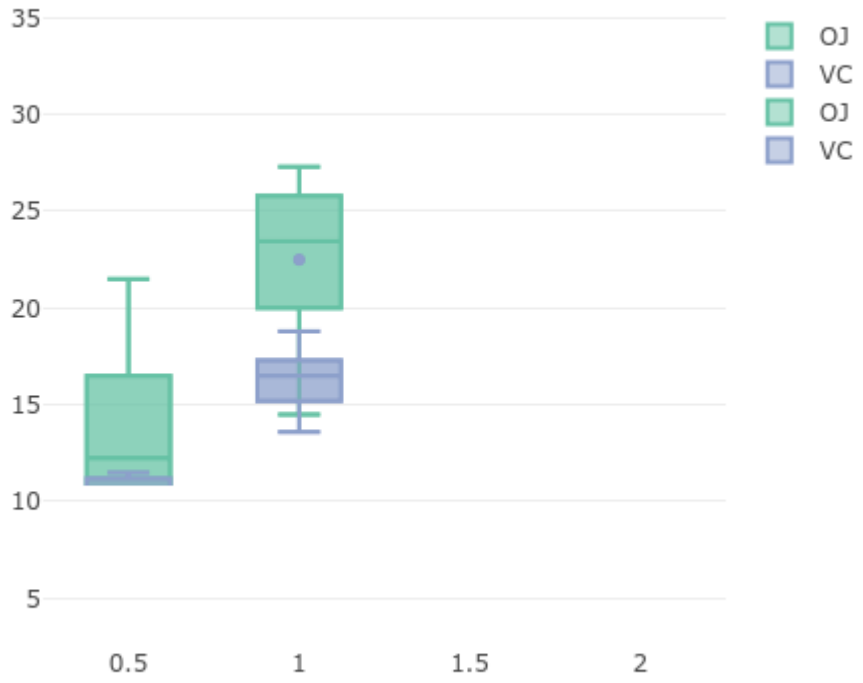
#Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose (Only use the techniques from class, even if there's other approaches worth considering)

```
# Set groups
supplement_vc_dose1<-filter(ToothGrowth, supp =="VC"&dose ==0.5)
supplement_vc_dose2<-filter(ToothGrowth, supp =="VC" & dose ==1)
vc_d3<-filter(ToothGrowth, supp =="VC" & dose ==2)
supplement_oj_dose1<-filter(ToothGrowth, supp =="OJ"& dose ==0.5)
supplement_oj_dose2 <-filter(ToothGrowth, supp =="OJ"& dose ==1)
```

```
oj_d3 <-filter(ToothGrowth, supp =="OJ"& dose ==2)
# Test the VC groups for alternative hypothesis
t.test (supplement_vc_dose1$len, supplement_vc_dose2$len)

##
##  Welch Two Sample t-test
##
## data:  supplement_vc_dose1$len and supplement_vc_dose2$len
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean of x mean of y
##      7.98     16.77

t.test (supplement_oj_dose1$len, supplement_oj_dose2$len)

##
##  Welch Two Sample t-test
##
## data:  supplement_oj_dose1$len and supplement_oj_dose2$len
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

The most commonly used way to visualize t-test-like comparison is to use boxplots. Below I provide the toothpaste example describing “relationship between tooth growth length and drinking supplement”

```
library(plotly)

#boxplot(len~supp*dose, data=tooth_data, notch=TRUE,
col=(c("red","gray")),
 # main="Tooth Growth", xlab="Suppliment and Dose")
plot_ly(ToothGrowth, x= tooth_data$dose, y= tooth_data$len, color =
tooth_data$supp, type = "box")%>%
  add_trace( x= tooth_data$dose, y= tooth_data$len)
```



Summary

Based on the data anasysis, lower dosages(0.5 - 1.0 mg) and orange juice provides more tooth growth than ascorbic acid. The higher dosage (2mg), the rate of tooth growth is not statistically different between supplement methods. Regardless of the supplement method, dosage is a key factor in tooth growth