Inferential Data Analysis

Ajith Sam

19/07/2020

Overview

In this project we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

2.Provide a basic summary of the data.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

4.State your conclusions and the assumptions needed for your conclusions.

Data Summary

```
library(datasets)
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
```

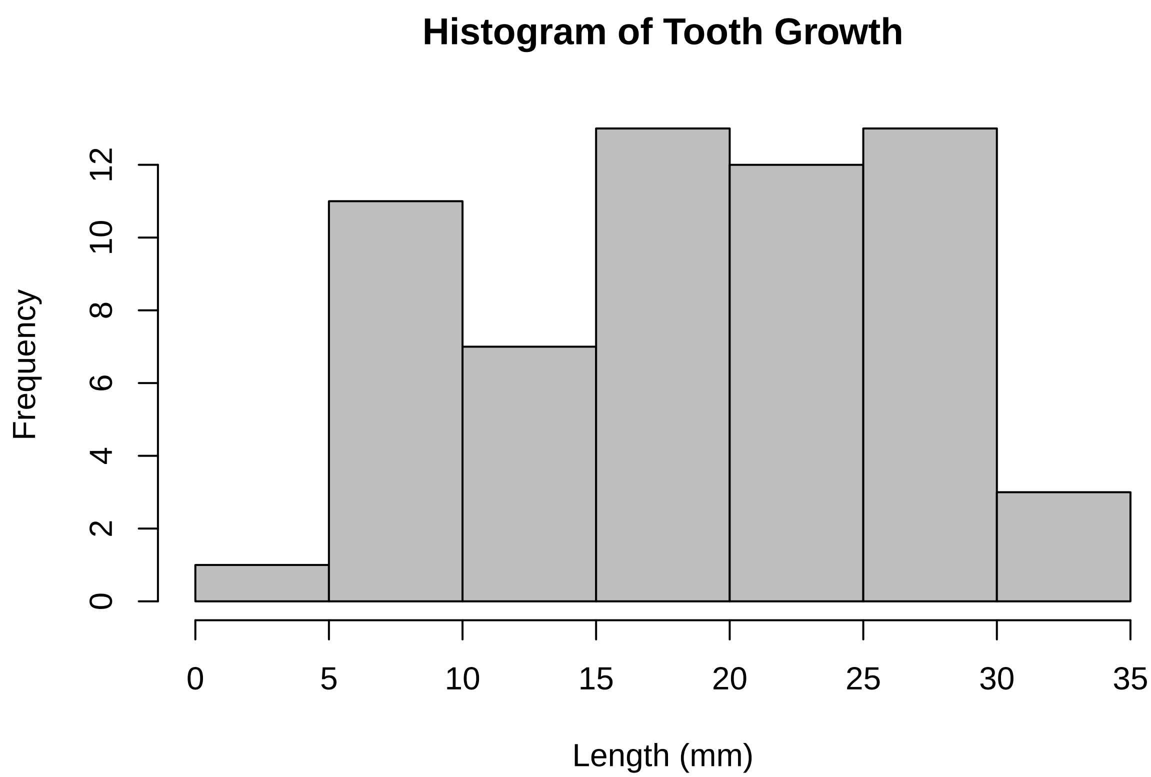
```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
#Variable "dose" conversion (numeric to factor):
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
#Data Visualization
hist(ToothGrowth$len, col = "gray",main = "Histogram of Tooth Growth",
     xlab = "Length (mm)",ylab = "Frequency")
```



Exploratory Data Analysis

```
summary(ToothGrowth)
```

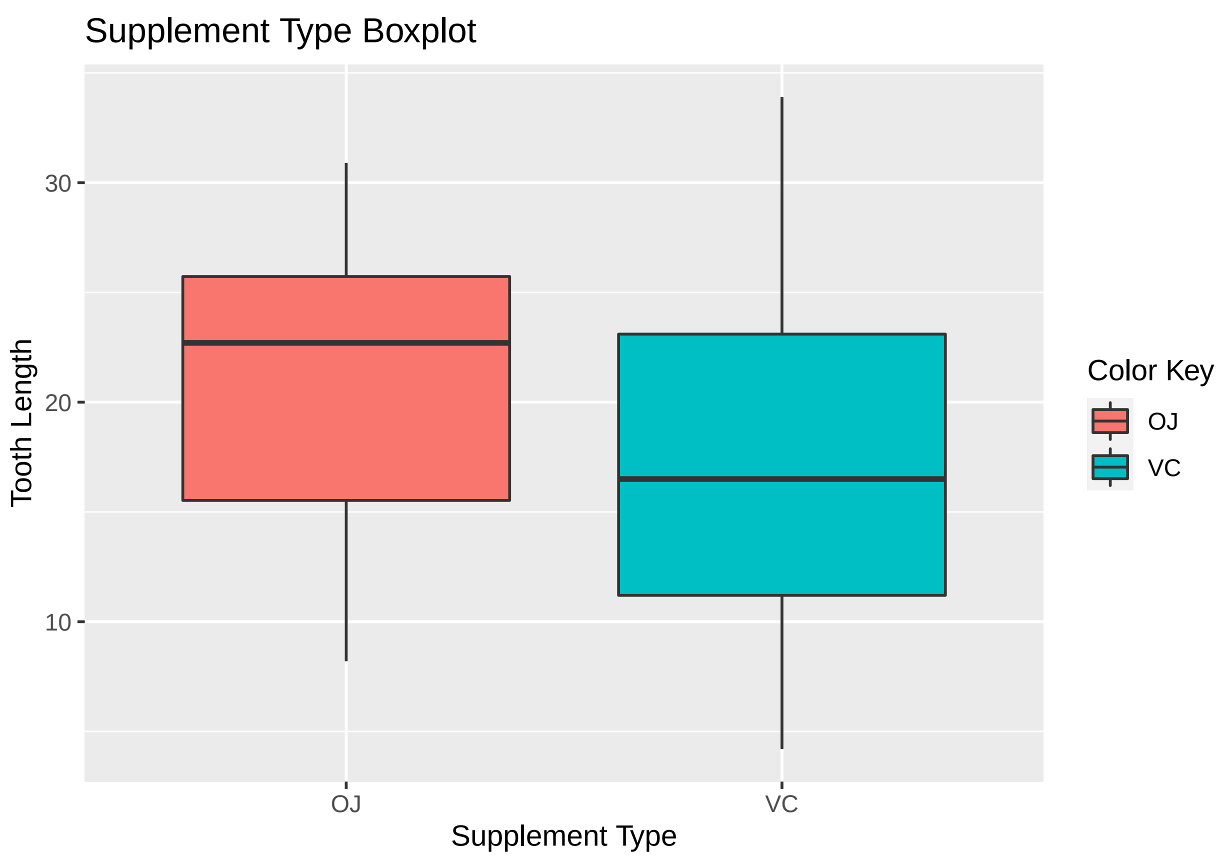
```
##       len        supp     dose   
##  Min.   : 4.20   OJ:30   0.5:20  
##  1st Qu.:13.07   VC:30   1  :20  
##  Median :19.25           2  :20  
##  Mean   :18.81                   
##  3rd Qu.:25.27                   
##  Max.   :33.90                   
```

```r
#Breakdown of supplement types vs dosage levels:
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```r
library(ggplot2)

#Visualization (Supplement type):
suppTyp <- ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
suppTyp <- suppTyp + labs(title = "Supplement Type Boxplot")
suppTyp <- suppTyp + xlab("Supplement Type")
suppTyp <- suppTyp + ylab("Tooth Length")
suppTyp <- suppTyp + labs(fill = "Color Key")
suppTyp
```



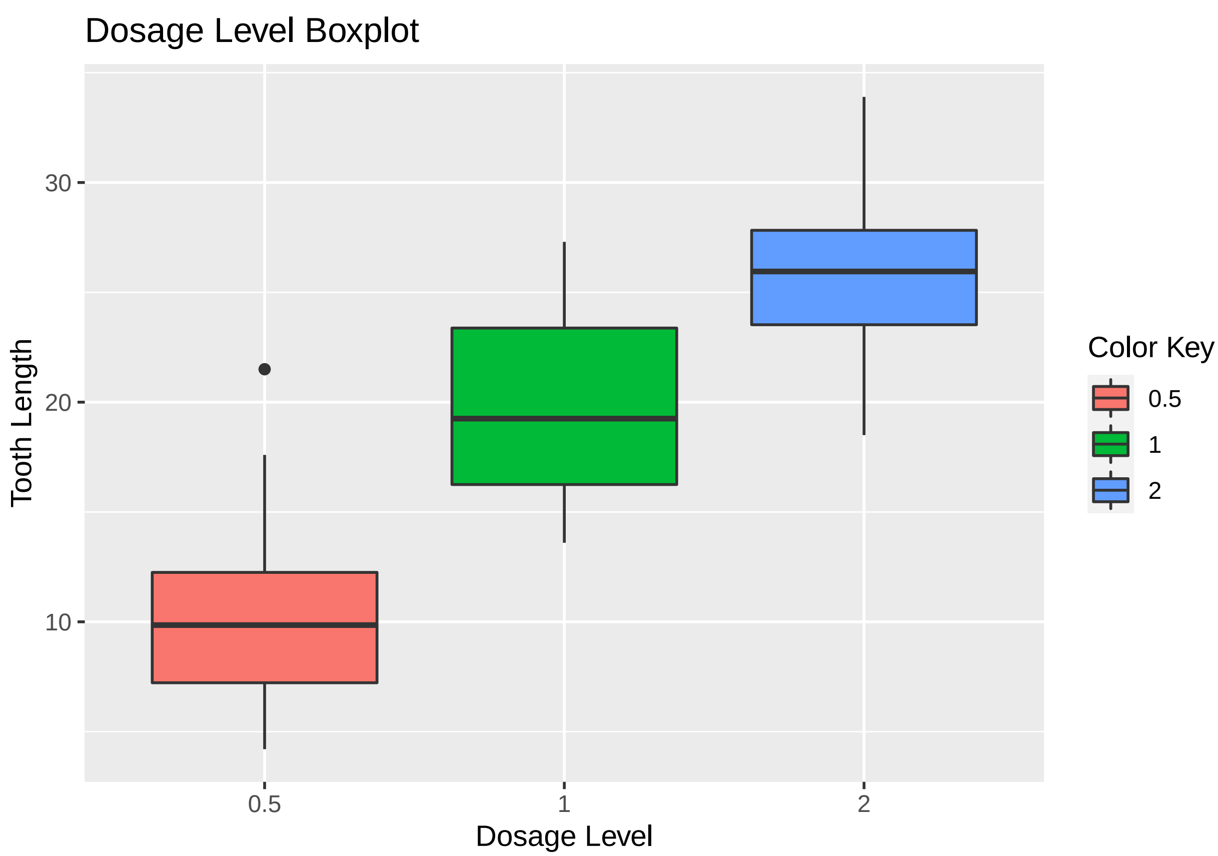

```r
#Supplement type boxplot mean and range:
suppTypMean = split(ToothGrowth$len, ToothGrowth$supp)
sapply(suppTypMean, mean)
```

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

```r
sapply(suppTypMean, range)
```

```
##        OJ   VC
## [1,]  8.2  4.2
## [2,] 30.9 33.9
```

```r
# Dosage level
dsg <- ggplot(aes(x = dose, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = dose))
dsg <- dsg + labs(title = "Dosage Level Boxplot")
dsg <- dsg + xlab("Dosage Level")
dsg <- dsg + ylab("Tooth Length")
dsg <- dsg + labs(fill = "Color Key")
dsg
```




```r
#Dosage level boxplot mean and range:
dsgMean = split(ToothGrowth$len, ToothGrowth$dose)
sapply(dsgMean, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```r
sapply(dsgMean, range)
```

```
##       0.5    1    2
## [1,]  4.2 13.6 18.5
## [2,] 21.5 27.3 33.9
```

SYNOPSIS : The statistical summary shows that average length of tooth growth over all the observations is 18.81mm between the range of 4.2mm and 33.9mm. We can also note from the statistical summary and breakdown of supplement types vs dosage levels that the amounts of each supplement and dosage were equally distributed across the 60 observations.

From the boxplots, mean, and range calculations presented above, the range of values for supplement types orange juice (OJ) and ascorbic acid (VC) are 8.2, 30.9, 4.2, 33.9mm respectively, denoting not much of a variance. Similarly, from the dosage level boxplot, we see increasing averages of 10.61, 19.73, 26.10mm for the 0.5, 1.0, & 2.0mg level dosages respectively as well as ever increasing range levels signifying a stronger factor relationship.

In summation, by comparing these two boxplots to each other, their respective means, and range levels one initially observes that dosage level factors more significantly than supplement type in overall tooth growth. In order to test this theory, hypothesis tests will be conducted below.

Hypothesis Test

Supplement Type

```r
#Hypothesis test to check factor significance of supplement type:
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

~ Fail to reject the null hypothesis, supplement delivery type is not correlated to tooth growth.

Dosage Level

```r
#Group subsetting prep for dosage level hypothesis tests:
dsgPt5vs1 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
dsgPt5vs2 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
dsg1vs2 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

```r
#Hypothesis test to check factor significance of dosage levels - 0.5 vs 1.0:
t.test(len ~ dose, data = dsgPt5vs1)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

~ Reject the null, accept the alternative hypothesis, dosage level 0.5 vs 1.0 is correlated to tooth growth.

```
#Hypothesis test to check factor significance of dosage levels - 0.5 vs 2.0:
t.test(len ~ dose, data = dsgPt5vs2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

~ Reject the null, accept the alternative hypothesis, dosage level 0.5 vs 2.0 is correlated to tooth growth.

```
#Hypothesis test to check factor significance of dosage levels - 1.0 vs 2.0:
t.test(len ~ dose, data = dsg1vs2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

~ Reject the null, accept the alternative hypothesis, dosage level 1.0 vs 2.0 is correlated to tooth growth.

Supplement Type and Dosage Level

```
#Group subsetting prep for supplement type and dosage level hypothesis tests:
dsgPt5 <- subset (ToothGrowth, dose == 0.5)
dsg1 <- subset (ToothGrowth, dose == 1.0)
dsg2 <- subset (ToothGrowth, dose == 2.0)
```

```r
#Hypothesis test to check factor significance of supplement type and dosage level 0.5:
t.test(len ~ supp, data = dsgPt5)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

~ Reject the null, accept the alternative hypothesis, supplement type and dosage level 0.5 is correlated to tooth growth.

```r
#Hypothesis test to check factor significance of supplement type and dosage level 1.0:
t.test(len ~ supp, data = dsg1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

~ Reject the null, accept the alternative hypothesis, supplement type and dosage level 1.0 is correlated to tooth growth.

```r
#Hypothesis test to check factor significance of supplement type and dosage level 2.0:
t.test(len ~ supp, data = dsg2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

~ Fail to reject the null hypothesis, supplement type and dosage level 2.0 is not correlated to tooth growth.

SYNOPSIS - Reject/do not reject hypothesis conclusions provided at the end of each hypothesis test above

Conclusion

Given the following assumptions:

1. The 60 guinea pigs observed were chosen at random
2. The 60 guinea pigs observed are a representative sample of the population
3. The supplement types and dosage levels were distributed equally and within a controlled environment, e.g. the same supplement type and dosage level combination, of the 6 variations, was provided to the same guinea pig initially and then repeatedly over the timeframe of the study

We conclude from the analysis that:

1. Initially, our exploratory data analysis indicated that supplement delivery type was not a significant factor in tooth growth
2. Both the exploratory data analysis and hypothesis tests concluded that dosage levels are a significant factor in tooth growth, with greater dosages providing for higher growth levels, all else equal
3. Finally, our initial observation that supplement delivery type is not a significant factor, was proven false. The combined supplement type and dosage levels of 0.5 and 1.0mg, respectively, are significant.
4. However, at the supplement type and dosage level of 2.0mg, it is not. I.e., supplement delivery type at the 0.5 and 1mg dosage levels factors significantly in tooth growth, while at higher levels (2mg), it does not.