Course Project (Part 2): Basic Inferential Data Analysis

Cesar Rene Pabon Bernal

July 24, 2016

Overview

In this report we aim to analyze the ToothGrowth data in the R datasets package by performing some basic exploratory data analyses. This will be accomplished by answering the following requests:

1. Load and explore the data
2. Provide a basic summary
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.

Load & Explore the Data

Let's first load the desired Toothgrowth package and its data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```
library(ggExtra)
```

```
## Warning: package 'ggExtra' was built under R version 3.2.5
```

```
library(datasets)
data(ToothGrowth)
```

Now let's explore the data by looking at

A. Its variables

```
## '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 ...
```

B. Number of of Rows

```
## ToothGrowth Data Number of Rows:  60
```

Basic Summary

A. Summary of all variables

```
##       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
```

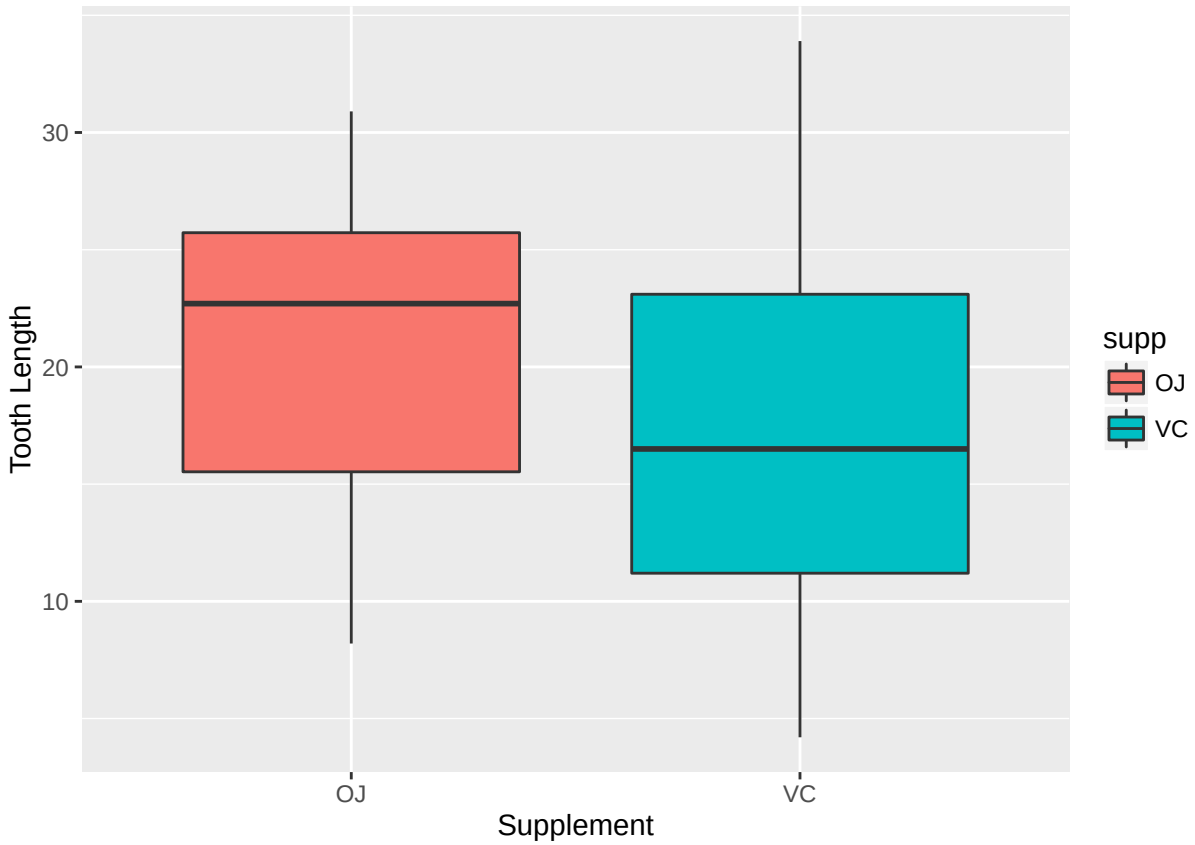
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

B. Summary of dose levels and delivery methods

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

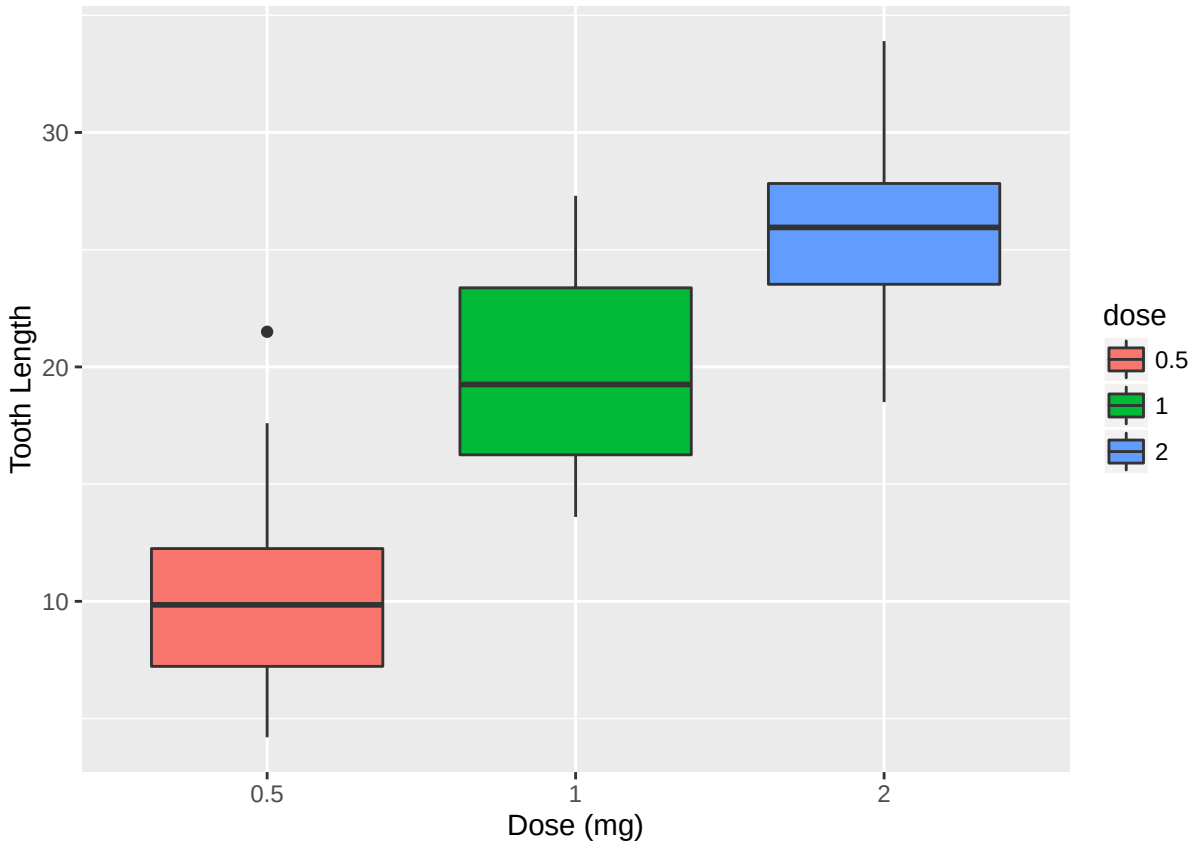
C. Box Plot of our summary between Supplement and Tooth Length demonstrating that OJ has a better effect on teeth growth than VC.

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
        xlab("Supplement") +ylab("Tooth Length")
```



D. Box Plot of Tooth Length relative to Dosage demonstrating there isn't much correlation between both variables.

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose)) +
        xlab("Dose (mg)") +ylab("Tooth Length")
```



Confidence Intervals and/or Hypothesis Tests

A. Lets first examine the group differences in accordance to different supplement types.

```
t.test(len ~ supp, paired = F, var.equal = F, 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
```

As we can see, the *p-value = 0.06063* and the confidence interval contains the value zero. Therefore we can't eliminate the hypothesis different supplement types have an effect on Tooth Length.

B. Now lets examine the data when Dosage is a Factor by analyzing the correlation between the Dosage and change in Tooth Growth.

```
dose_1 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose_2 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose_3 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose_1)
```

```
##
##  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
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose_2)
```

```
##
##  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
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose_3)
```

```
##
##  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
```

C. Lastly, lets examine the data when Supplement is a Factor (within Dose Levels) by analyzing the correlation between Dosage and Tooth Growth

```
dose_4 <- subset(ToothGrowth, dose == 0.5)
dose_5 <- subset(ToothGrowth, dose == 1.0)
dose_6 <- subset(ToothGrowth, dose == 2.0)
t.test(len ~ supp, paired = F, var.equal = F, data = dose_4)
```

```
##
##  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
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = dose_5)
```

```
##
##  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
```

```
t.test(len ~ supp, paired = F, var.equal = F, data = dose_4)
```

```
##
##  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
```

As we can see, all three of the *p-values* < 0.05 (significantly) and the confidence interval does not contain zero. In addition, the *mean* Tooth Length increases when Dosage increases and thus, we are unable to eliminate the hypothesis.

Conclusions and Assumptions

In conclusion, we have discovered that increasing the Dosage leads to an increase in Tooth Growth and that types of Supplements used have no effect on Tooth Growth.

Based on the information, we can assume that the experiment was conducted in a generalized manner in which the conclusions can hold validity. Details such as randomized selection of subjects, Dosage, and Supplements, hold positive to that claim.