# PART 2: Basic Inferential Data Analysis

WendyD

9/9/2020

```
library(ggplot2)
```

## Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
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 ...
```

```
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
```

## Basic Summary of the data

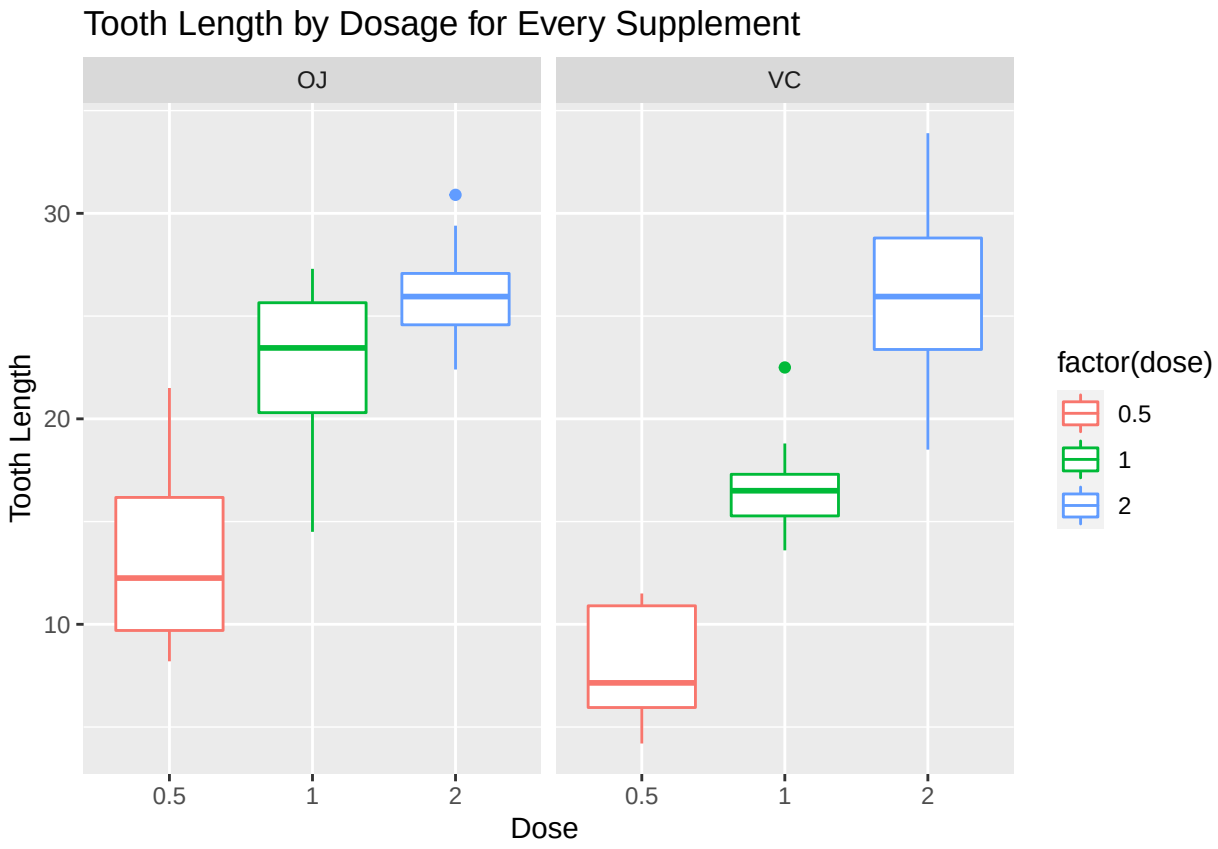
```
summary(ToothGrowth)
```

```
##       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
```

Since our data has 3 variables: len(tooth length), supp (supplement types), and dose (in mg/day), we could explore our data, using dosage as factor variable, and to see how the tooth length changes with different dosage accordingly to their supplement types

```
ggplot(data=ToothGrowth, aes(x=factor(dose), y=len, colour=factor(dose)))+
               facet_grid(.~ToothGrowth$supp)+
               geom_boxplot()+
```

```
        xlab("Dose ")+ ylab("Tooth Length")+
        ggtitle("Tooth Length by Dosage for Every Supplement")
```




From the graph, it appreared that Orange juice outperformed ascorbic acid at dose 0.5mg/day and 1mg/day We can also compare the effectivty of 2 supplements by calculating the mean by supplements factor

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

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

This confirm our observation, Orange juice does increase the tooth length faster than ascirbic acid

# Hypothesis Testing to compare tooth growth by supp and dose

## Analyze tooth growth by supplement

We test two hypothesis: First: there is no correlation in the tooth length for different supplements. Secondly, there is no correlation in tooth length acorss different doses of supplement.

Tooth length for those taking orange juice as supplement

```
orangejuice<- subset(ToothGrowth
                     $len, ToothGrowth$supp == "OJ")
```

Tooth length for those taking ascorbic acid as supplement

```
aacid<- subset(ToothGrowth$len, ToothGrowth$supp =="VC")
```

Compute t test to see if there is any significant difference in the tooth length for people taking different supplement

```
t.test(orangejuice, aacid )
```

```
##
##  Welch Two Sample t-test
##
## data:  orangejuice and aacid
## 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 of x mean of y
##  20.66333  16.96333
```

Since the p value is $>$ than 0.05, we cannot reject the null hypothesis, and concluded that there is no significant difference in tooth length for different supplements

## Analyze tooth growth across dosages

```
dose0.5<- subset(ToothGrowth$len, ToothGrowth$dose == "0.5")
dose1<- subset(ToothGrowth$len, ToothGrowth$dose == "1")
dose2<- subset(ToothGrowth$len, ToothGrowth$dose == "2")
```

Compute t test to see if there is any significant difference in the tooth length for people taking different dosesages

Dose 0.5 and dose 1mg/day

```
t.test(dose0.5, dose1,paired = FALSE, alternative = "less",  var.equal = FALSE )
```

```
##
##  Welch Two Sample t-test
##
## data:  dose0.5 and dose1
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Dose 1 and dose 2

```
t.test(dose1, dose2,paired = FALSE, alternative = "less",  var.equal = FALSE )
```

```
##
##  Welch Two Sample t-test
##
## data:  dose1 and dose2
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
```

```
## mean of x mean of y
##    19.735    26.100
```

Dose 0.5 and Dose 2 mg/day

```
t.test(dose0.5, dose2,paired = FALSE, alternative = "less",  var.equal = FALSE )
```

```
##
##  Welch Two Sample t-test
##
## data:  dose0.5 and dose2
## t = -11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.27926
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Since the p value for both 3 t test are very small, we have to reject the null hypothesis and concluded that there are correlation in dosages and tooth length

## Conclusion

We concluded that there is no significant correlation in taking orange juice or ascorbic acid in regard to tooth lengths. We analyzed the correlation between 0.5 and 1 mg/day, 0.5 and 2mg/day, and 1mg and 2mg/day. There are strong correlation between 3 different dosages and the tooth length to one another.