# ToothGrowth_Basic Inferential Data Analysis

*Yato*

*March 16, 2017*

## Overview

This part is going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data, and provide a basic summary of the data.
2. Perform some basic exploratory data analyses.
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.

## Loading the data and exploratory data analyses

```
data(ToothGrowth); attach(ToothGrowth)
```

## Basic Summary of the Data

```
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 ...
```

```
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
```

```
unique(dose)
```

```
## [1] 0.5 1.0 2.0
```

```
table(supp, dose)
```

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

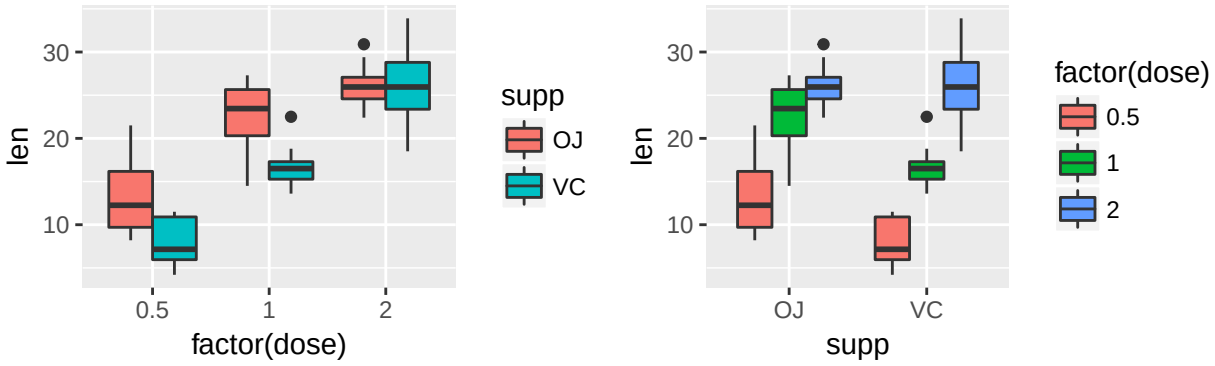
```
ToothGrowth %>% group_by(supp, factor(dose)) %>% summarise(mean_len = mean(len), sd_len = sd(len))
```

```
## Source: local data frame [6 x 4]
## Groups: supp [?]
##
##     supp `factor(dose)` mean_len   sd_len
```

```
##   <fctr>         <fctr>    <dbl>    <dbl>
## 1     OJ            0.5    13.23 4.459709
## 2     OJ              1    22.70 3.910953
## 3     OJ              2    26.06 2.655058
## 4     VC            0.5     7.98 2.746634
## 5     VC              1    16.77 2.515309
## 6     VC              2    26.14 4.797731
```

## Exploratory Data Analysis

```
gg1<-ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = supp)) + geom_boxplot()
gg2 <- ggplot(ToothGrowth, aes(x = supp, y = len, fill = factor(dose))) + geom_boxplot()
grid.arrange(gg1, gg2, ncol = 2) # or interaction2wt(len ~ supp*factor(dose))
```




## Confidence Interval and Hypothesis Testing

- Examining the impact of **supp** or **dose** as a factor separately.
- Examining the impact of supp, dose and the interaction of supp and dose by including both **supp and dose** as factors.

```
dose15 <- ToothGrowth %>% filter(dose %in% c("0.5", "1"))
dose12 <- ToothGrowth %>% filter(dose %in% c("1", "2"))
dose52 <- ToothGrowth %>% filter(dose %in% c("0.5", "2"))
t.test(len ~ supp)
```

```
##
##  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
```

```
t.test(len ~ factor(dose), data = dose15)
```

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

```
##
## data:  len by factor(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 ~ factor(dose), data = dose12)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(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
```

```
t.test(len ~ factor(dose), data = dose52)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(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
```

```
# or summary(aov(len ~ dose)) and summary(aov(len ~ supp*factor(dose)))
detach(ToothGrowth)
```

## Conclusion

- This is a balanced design(equal sample sizes in each cell of the design).
- The confidence interval is [-0.171, 7.571], which include zero. There is no significant correlation between delivery methods and tooth length.
- The confidence interval is [-11.98, -6.276] for doses 0.5 and 1.0, [-18.16, -12.83] for doses 0.5 and 2.0, and [-8.996, -3.734] for doses 1.0 and 2.0) and show that there is a significant correlation between tooth length and dose levels.
- T-test shows that two delivery methods (OJ or VC) have equal tooth length ($p>0.05$). Three levels of dose aren't all equally effective. Increasing the dose level can lead to increased tooth growth.
- Both main effects and the interaction between these factors are significant.