# ToothGrowth Inferential Data Analysis

*Charles Yoo*

*June 19, 2016*

## Overview

This project investigages the correlation that various dosages and supplements have on tooth length using the ToothGrowth dataset. The null hypothesis is that OJ has a stronger efficacy on tooth growth than Vitamin C.

## Load Data

```
data(ToothGrowth)
```

## Summary of Data

```
# description of variables
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 of dataset
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
```

```
# look at first 6 rows
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
```

```r
# count number of rows
nrow(ToothGrowth)
```

```
## [1] 60
```

```r
dim(ToothGrowth)
```

```
## [1] 60  3
```

```r
# exploration of data
unique(ToothGrowth$dose)
```

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

```r
unique(ToothGrowth$supp)
```

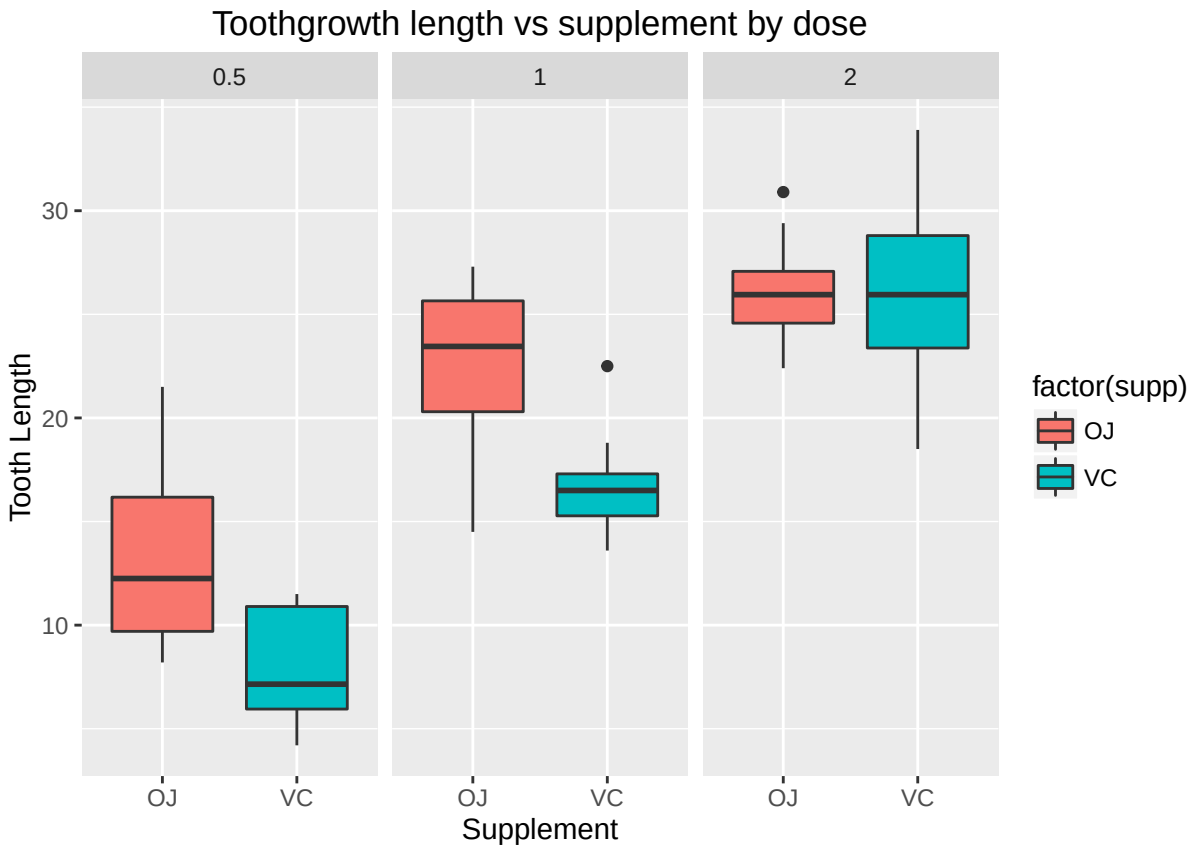
```
## [1] VC OJ
## Levels: OJ VC
```

```r
# table comparing dose to supp
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

## Comparison of dose to supp

```r
ggplot(ToothGrowth, aes(x = supp, y = len)) +
        geom_boxplot(aes(fill = factor(supp))) +
        facet_grid(. ~ dose) +
        xlab("Supplement") +
        ylab("Tooth Length") +
        ggtitle("Toothgrowth length vs supplement by dose")
```




We see a visual correlation between dose size and tooth length. The higher the dosage of the supplement affects a longer tooth length. We also see that the type of supplement may have an effect on tooth length as well at the lower dosages.

## Confidence intervals

*Null-hypotheis*

OJ supplement has a more impact than Vitamin C.

### T-test supplment as factor

Assign variables

```
length <- ToothGrowth$len
supplement <- ToothGrowth$supp
dosage <- ToothGrowth$dose
```

T-test on supplement

```
t.test(length[supplement == "OJ"], length[supplement == "VC"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  length[supplement == "OJ"] and length[supplement == "VC"]
```

```
## 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
```

The confidence interval [-0.1710156, 7.5710156] and the p value 0.06063 shows a significance. The difference in mean is not significant enough to support the null hypothesis.

### T-test dosages as factor

Assign variables

```
dose0510 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose0520 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose1020 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

T-test on doses 0.5 & 1.0

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose0510)
```

```
##
##  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
```

The confidence interval [-11.983781, -6.276219] and the p-value 1.268e-07, essentially 0, shows a significant correlation between tooth length and dose levels. There is a significant difference between the dosages that the null hypothesis is true.

0.5 & 2.0

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose0520)
```

```
##
##  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
```

The confidence interval [-18.15617, -12.83383] and the p-value 4.398e-14, essentially 0, shows a significant correlation between tooth length and dose levels. There is a significant difference between the dosages that the null hypothesis is true.

1.0 & 2.0

```
t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = dose1020)
```

```
##
##  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
```

The confidence interval [-8.996481, -3.733519] and the p-value 1.906e-05, essentially 0, shows a significant correlation between tooth length and dose levels. There is a significant difference between the dosages that the null hypothesis is true.

### T-test supplement as factor within dosage levels

Assign variables

```
dose05 <- subset(ToothGrowth, dose == 0.5)
dose10 <- subset(ToothGrowth, dose == 1.0)
dose20 <- subset(ToothGrowth, dose = 2.0)
```

T-test on supplement with doses 0.5

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose05)
```

```
##
##  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
```

The confidence interval [1.719057, 8.780943] and the p-value 0.006359, essentially 0, shows a significant correlation between tooth length and dose levels. There is a significant difference between the dosages that the null hypothesis is true.

1.0

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose10)
```

```
##
##  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
```

The confidence interval [2.802148, 9.057852] and the p-value 0.001038, essentially 0, shows a significant correlation between tooth length and dose levels. There is a significant difference between the dosages that the null hypothesis is true.

2.0

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = dose20)
```

```
##
##  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
```

The confidence interval [-0.1710156 7.5710156] and the p-value 0.06063, essentially 0, shows a significant correlation between tooth length and dose levels. The difference in mean is not significant enough to support the null hypothesis.

## Conclusions

- There is no significance difference in tooth lengths when looking at supplements only.
- There is a significance difference in tooth lengths when looking at dosages only.

### References

http://rstudio-pubs-static.s3.amazonaws.com/56907_9a9016c4f4df4d10aa7afaed5eed1d77.html
https://rpubs.com/YevgenyY/si_cp2
http://rstudio-pubs-static.s3.amazonaws.com/26135_b10068633b5948848d2c28fc0a752d78.html