# Statistical Inference Course Project Part 2

*Tulin Varol*

*04/06/2018*

## Overview

In this project, the Tooth Growth dataset from R dataset package will be studied. The decscription of the dataset can be found here. First, basic exploratory analysis will be performed in order to understand the data better. Then, the hypothesis testing and confidence interval calculations will be performed to compare tooth growth by supplement type and dosage. Finally, the conclusions and assumptions made will be given.

## A Quick Look at the Tooth Growth Dataset

```
# Load the ToothGrowth data
dataset <- ToothGrowth

# Print the summary of the data
summary(dataset)
```

```
##       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  
```

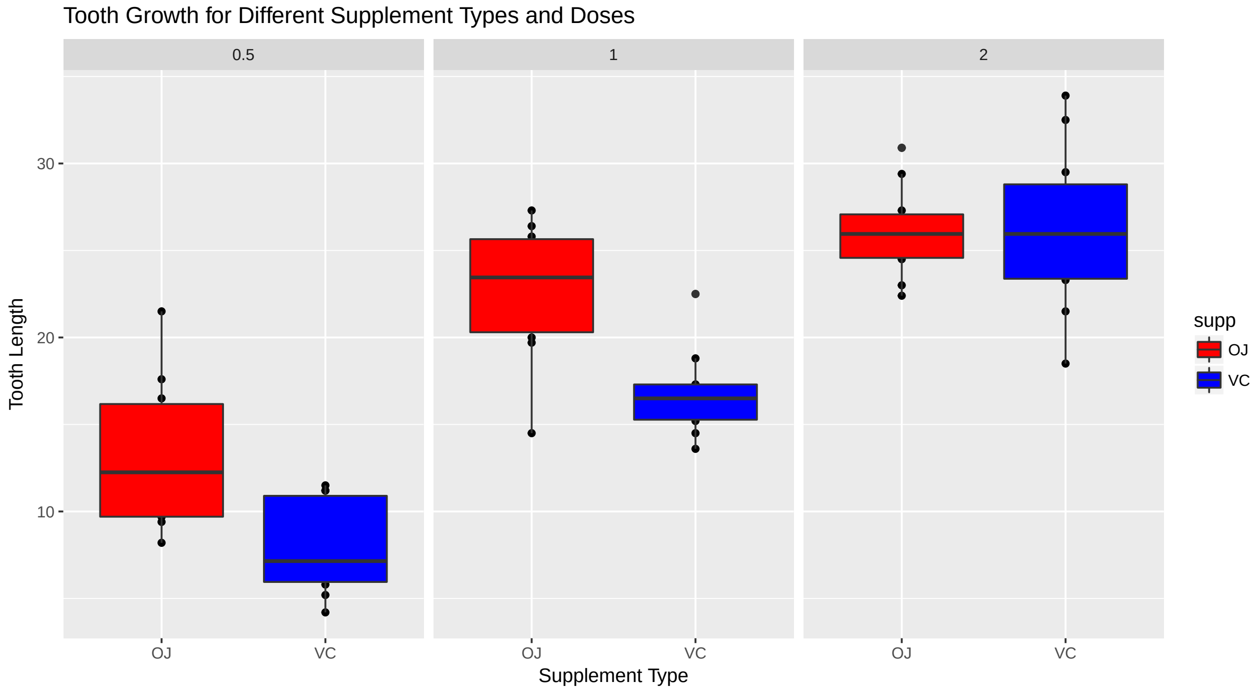
A few highlights from the dataset:

- There are 60 observations and 3 variables
- Tooth length varies from 4.2 to 33.9 with an average of 18.8 (probably in mm)
- There are two different supplement types: OJ (orange juice), VC (vitamin C)
- There are three dose levels: 0.5, 1, 2 mg/day

## Basic Exploratory Data Analysis

Let's have a quick look at the data to see how tooth length varies with the supplement type and dose.

```
library(ggplot2)
qplot(supp,len,data=dataset, facets=~dose,
      main="Tooth Growth for Different Supplement Types and Doses",
      xlab="Supplement Type",
      ylab="Tooth Length") + geom_boxplot(aes(fill = supp)) +
    scale_fill_manual(values=c("red", "blue"))
```




It seems that, when dosage increases, tooth length also increases. Except for 2 mg/day case, OJ supplement has a bigger effect on tooth length comparing to the supplement VC. However, for the dosage of 2 mg/day, medians of the measurements from both supplements are almost the same.

# Hypothesis Tests

Let's first perform a hypothesis test to see if the supplement type really has an affect on tooth growth when all doses are combined. We want to test the null hypothesis "The supplement type has no impact on tooth growth" against the alternative hypothesis "The supplement type has an impact on tooth growth".

```
supp.test.all <- t.test(len~supp, paired = FALSE,  var.equal = FALSE, data = dataset)
supp.test.all$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
supp.test.all$p.value
```

```
## [1] 0.06063451
```

When there is no impact of the supplement type on tooth growth, one expects to see a difference of zero between means of tooth growth measurement from two different supplements. From the results of t-test, 95% confidence intervals include zero, meaning that one fails to reject the null hypothesis which claims "there is no dependence on the supplement type". This can also be concluded from p-value since it is higher than 0.05.

One can perform this hypothesis test for different dosage of supplements. The hypothesis tests performed for this can be found in Appendix. The conclusion from those tests is that tooth growth has a dependence on the supplement type for lower dosage (0.5 and 1 mg/day) but this is not the case when dosage is 2 mg/day.

Now let's check the dependence of tooth growth on the dosage of supplements. Here we'll test the null hypothesis of "Different dosage of supplements has no impact on tooth growth" against the alternative hypothesis of "Different dosage of supplements has an impact on tooth growth". In order to do this, we choose the extreme dosage in our dataset; 0.5-2 mg/day

```
# Consider OJ, and check the affect of dosage
low.dosage.OJ = dataset$len[dataset$supp=="OJ" & dataset$dose==0.5]
high.dosage.OJ = dataset$len[dataset$supp=="OJ" & dataset$dose==2]
dose.test.OJ <- t.test(low.dosage.OJ, high.dosage.OJ, paired = FALSE, var.equal = FALSE)
dose.test.OJ$conf.int
```

```
## [1] -16.335241  -9.324759
## attr(,"conf.level")
## [1] 0.95
```

```
dose.test.OJ$p.value
```

```
## [1] 1.323784e-06
```

```
# Consider VC, and check the affect of dosage
low.dosage.VC = dataset$len[dataset$supp=="VC" & dataset$dose==0.5]
high.dosage.VC = dataset$len[dataset$supp=="VC" & dataset$dose==2]
dose.test.VC <- t.test(low.dosage.VC, high.dosage.VC, paired = FALSE, var.equal = FALSE)
dose.test.VC$conf.int
```

```
## [1] -21.90151 -14.41849
## attr(,"conf.level")
## [1] 0.95
```

```
dose.test.VC$p.value
```

```
## [1] 4.681577e-08
```

Indepent of the supplement type, the alternative hypothesis is favored strongly. This can be concluded from very small p-values and confidence intervals not including zero. Thus, tooth length has a dependence on dosage.

# Assumptions

- We perform an unpaired test since test subjects are not related, and are randomly assigned supplements/dose (paired = FALSE taken in t.test).
- For different groups, we assume different variances, not a constant one (var.equal = FALSE taken in t.test).

# Conclusions

### Conclusions from the Exploratory Analysis

- OJ supplement has a bigger effect on tooth length comparing to the supplement VC in all cases except for the case when the dose given is 2 mg/day.
- An increase in dosage results in an increase in tooth length.

### Conclusions from Hypothesis Tests

- When considering all doses combined, there is no dependence of tooth growth on a supplement type.
- When grouping subjects depending on the dosage of supplements given, a dependence of tooth growth on a supplement type is observed for dosage of 0.5 and 1 mg/day, whereas for a dosage of 2 mg/day, no such dependence is observed.
- Indepent of the supplement type, it is observed that different dosage of supplements has an impact on tooth growth.

# APPENDIX

```r
t.test(len~supp, paired = FALSE, var.equal = FALSE, data = subset(dataset, dose==0.5))
```

```
##
##  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
```

```r
t.test(len~supp, paired = FALSE, var.equal = FALSE, data = subset(dataset, dose==1))
```

```
##
##  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
```

```r
t.test(len~supp, paired = FALSE, var.equal = FALSE, data = subset(dataset, dose==2))
```

```
##
##  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
```

That's interesting! We actually observe differences on tooth growth for different supplements when supplement dose is 0.5 or 1 mg/day. In these two cases, confidence intervals don't include zero and p-values are smaller than 0.05, thus, the alternative hypothesis is favored. However, when supplement dose is 2 mg/day, we fail to reject the null hypothesis since the confidence interval includes zero and p-value is greater than 0.05. Therefore, one can conclude that tooth growth has a dependence on the supplement type for lower dosage values but this is not the case when dosage is 2 mg/day.