# Statistical Inference Course Project: The ToothGrowth data analysis

*Michela Ieva*

*Friday, May 22, 2015*

## Abstract

We present the results of a study about the *ToothGrowth* dataset. Data describe the teeth growth in guinea pigs as result of different type of treatments with Vitamin C. Differences are in dosage: 0.5, 1 and 2 mg and type of supplement: orange juice or ascorbic acid. The goal is understand if the treatments were effective.

```
library(knitr)
library(ggplot2)

echo = TRUE  # Always make code visible
```

## Data sample overview

We download the dataset from R library.

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

First of all we need to take a look at data format and 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 ...
```

So, as you can see data are organized in a data frame of 60 observations and 3 variables about the teeth lenght, the type of supplement (OJ or VC) and the dosis.

Here a statistic summary of data sample:

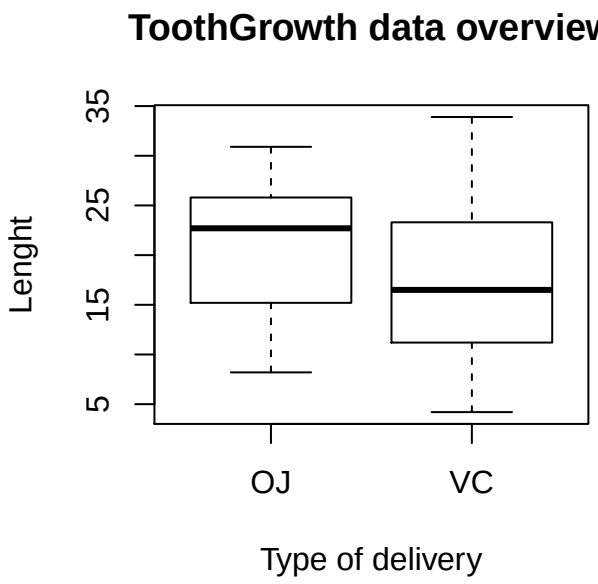
```
summary(ToothGrowth)
```

```
##       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
```

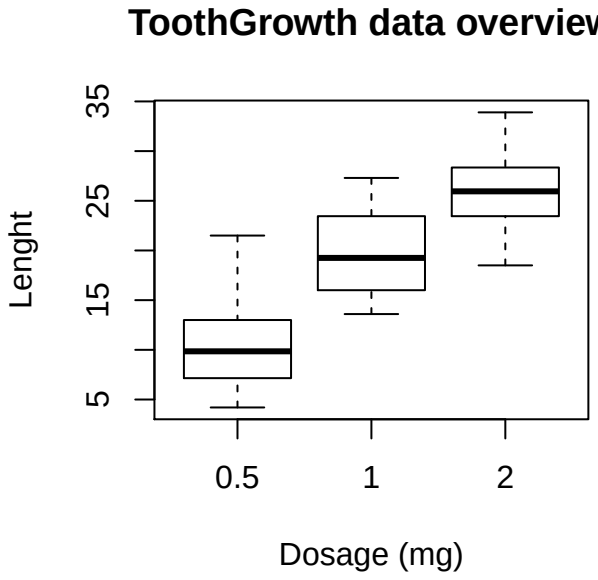
### Data preliminary analysis

To understand the effects of the treatments we perform some basic plot:

```
attach(ToothGrowth)
plot(supp, len, main = "ToothGrowth data overview", xlab = "Type of delivery", ylab = "Lenght")
```



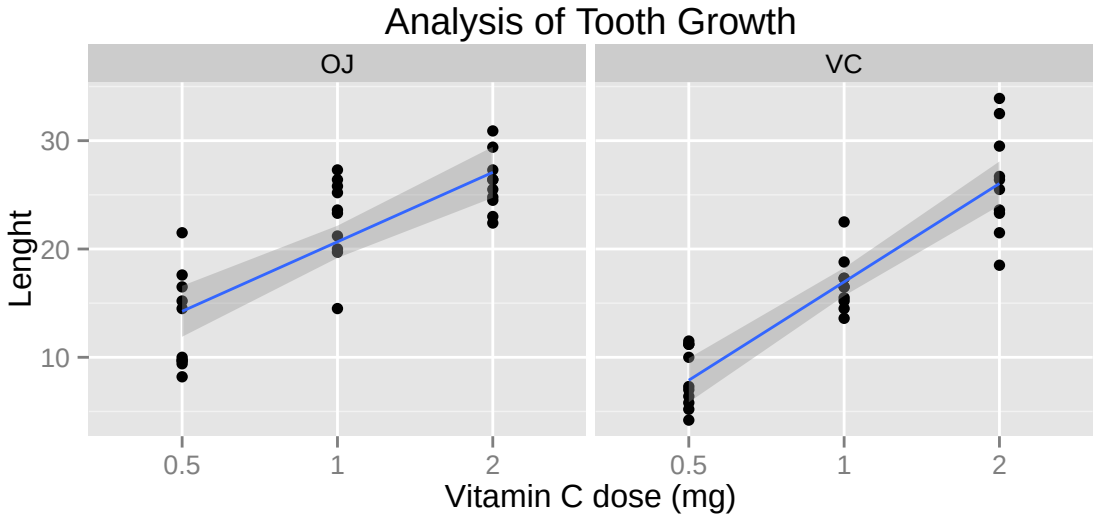

```
plot(factor(dose), len, main = "ToothGrowth data overview", xlab = "Dosage (mg)", ylab = "Lenght")
```



It seems, from the first plot, orange juice should be a little bit more effective than ascorbic acid that shows, however, a bigger variability. Second plot shows a clear trend in increasing lenght with higher dosage.

It is better try to visualize both effects in a more quantitative way, as follow:

```
ggplot(data=ToothGrowth, aes(x=factor(dose), y=len))+geom_point() + geom_smooth(method = "lm", aes(grou
```




Here the increasing of lenght as function of dosage is clear also from the fit, instead the effectiveness of a particolar type of supplement is not so evident.

## Statistical analysis : CI and hypothesis tests

To be more quantitive and accept or reject the two null hypothesis:

1. There is no difference on tooth length respect to supplement type(OJ and VC);
2. There is no difference on tooth length respect to different dosage.

we calculate a CI intervals and hypothesis tests on datasample.

First of all, we need to subset data into smaller sample by *dose* and *supp* variables to disentangle effects.

```
sampleOJ_1<-subset(ToothGrowth, (dose==0.5 & supp == "OJ"))
sampleOJ_2<-subset(ToothGrowth, (dose==1.0 & supp == "OJ"))
sampleOJ_3<-subset(ToothGrowth, (dose==2.0 & supp == "OJ"))

sampleVC_1<-subset(ToothGrowth, (dose==0.5 & supp == "VC"))
sampleVC_2<-subset(ToothGrowth, (dose==1.0 & supp == "VC"))
sampleVC_3<-subset(ToothGrowth, (dose==2.0 & supp == "VC"))
```

Now each single sample looks like as follow,

```
sampleOJ_1
```

```
##     len supp dose
## 31 15.2   OJ  0.5
## 32 21.5   OJ  0.5
## 33 17.6   OJ  0.5
## 34  9.7   OJ  0.5
## 35 14.5   OJ  0.5
## 36 10.0   OJ  0.5
## 37  8.2   OJ  0.5
## 38  9.4   OJ  0.5
## 39 16.5   OJ  0.5
## 40  9.7   OJ  0.5
```

To analyze data and accept or reject the null hypothesis, we have to use t-confidence interval, instead of normal, because of the smallness of the sample and we should suppose the samples are independent because there are no different info about that and furthermore with different variances.

So, to check the effectiveness of different dosage we compare the sample 1, 2, and 3 for the same type of supplement:

```
t.test(sampleOJ_2$len, sampleOJ_1$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  5.524366 13.415634
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(sampleOJ_3$len, sampleOJ_1$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```
t.test(sampleOJ_3$len, sampleOJ_2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 0.1885575 6.5314425
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(sampleVC_2$len, sampleVC_1$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  6.314288 11.265712
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(sampleVC_3$len, sampleVC_1$len, paired = FALSE, var.equal = FALSE)$conf
```

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

```
t.test(sampleVC_3$len, sampleVC_2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  5.685733 13.054267
## attr(,"conf.level")
## [1] 0.95
```

On the contrary to test the hypothesis of different effect related to the use of orange juice or ascorbic acid, we should compare same dosage groups but different type of supplements.

```
t.test(sampleOJ_1$len, sampleVC_1$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(sampleOJ_2$len, sampleVC_2$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(sampleOJ_3$len, sampleOJ_3$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -2.494589  2.494589
## attr(,"conf.level")
## [1] 0.95
```

## Conclusions

So, about the first null hypothesis:

1. There is no difference on tooth length respect to supplement type(OJ and VC).

According the results of t-confidence test at 95% CI, the first two intervals relative to dosages 0.5 and 1 mg are completely above zero than we can reject the null hypothesis and conclude a different effect on the teeth growth. Different the case of the last dosage, that contains zero-value, so we fail to reject the null hypothesis and conclude that in this case the effect of the dosage overcome the type of supplement.

About the second null hypothesis:

2. There is no difference on tooth length respect to different dosage.

According the results of t-confidence test at 95% CI, all the intervals comparing different dosages lie completely above zero for both type of supplements. so we can reject the null hypothesis and conclude the strong effect of different tretement dosage on the teeth growth, as shown also by the clear trend in the initial plots.