

# Tooth Growth

## Main goal

Analyze the ToothGrowth dataset. Perform a basic inferential data analysis.

## Exploratory analysis

### Load the dataset

```
data(ToothGrowth)
summary(ToothGrowth)
```

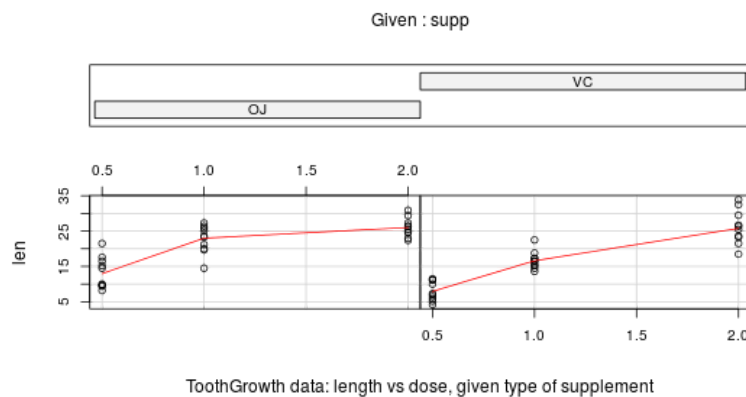
```
##      len      supp      dose
## Min.   : 4.2    OJ:30    Min.   :0.50
## 1st Qu.:13.1   VC:30    1st Qu.:0.50
## Median :19.2             Median :1.00
## Mean   :18.8             Mean   :1.17
## 3rd Qu.:25.3             3rd Qu.:2.00
## Max.   :33.9             Max.   :2.00
```

The dataset contains 60 observations about the effect of vitamin C on tooth growth in guinea pigs: *len* is the dependent variable, *supp* and *dose* are the independent variables.

### Plot the dataset

This plot is provided by the R package:

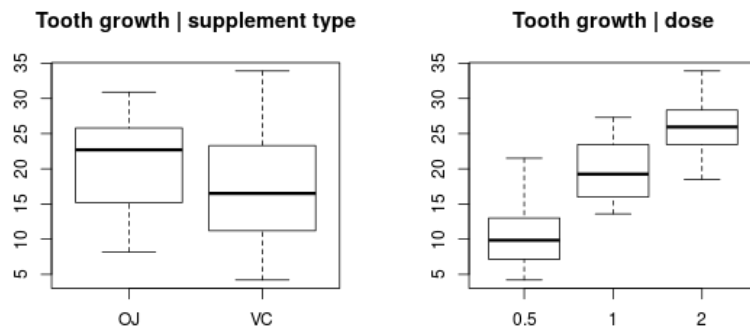
```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



It shows that there is a clear link between higher doses and tooth length. The link between supplement type is not so clear, i.e. to the naked eye, the curves look pretty similar for both supplements.

We can also plot boxes for each variable: first, tooth length as a function of the supplement type, then tooth length as a function of the dose.

```
par(mfrow=c(1,2))
boxplot(len ~ supp, data = ToothGrowth, main = "Tooth growth | supplement type")
boxplot(len ~ dose, data = ToothGrowth, main = "Tooth growth | dose")
```



Again, the link between dose and tooth growth looks pretty clear.

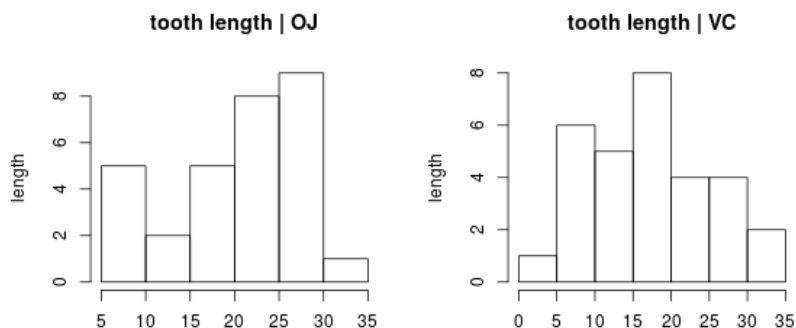
## Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

We are going to compare samples (two at a time) using *Student's t-test*.

### Assumptions:

- 1. the two samples are taken from populations that follow a Gaussian distribution. Histograms for tooth length show that they are indeed quite normal, if a bit skewed:

```
par(mfrow=c(1,2))
hist(ToothGrowth[ToothGrowth$supp == 'OJ',][1]$len, main = "tooth length | OJ", ylab = "length", xlab = "")
hist(ToothGrowth[ToothGrowth$supp == 'VC',][1]$len, main = "tooth length | VC", ylab = "length", xlab = "")
```



- 2. the samples are *iid*.

Only if there is a direct relationship between each specific pig in each sample, then can we set the *paired* parameter to TRUE. I found the dataset introduction unclear, but later learned that the original paper by Crampton, E.W. makes it clear that these observations come from 60 different guinea pigs, as each had to be killed to take the measurement :(

- 3. the number of observations is rather small, so Student's t-test is appropriate.

### Test Setup

We divide the observations in arrays by (dose + supplement type)

```
dose05_suppOJ <- ToothGrowth[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 0.5,]$len
dose05_suppVC <- ToothGrowth[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 0.5,]$len

dose1_suppOJ <- ToothGrowth[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 1,]$len
dose1_suppVC <- ToothGrowth[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 1,]$len

dose2_suppOJ <- ToothGrowth[ToothGrowth$supp == 'OJ' & ToothGrowth$dose == 2,]$len
dose2_suppVC <- ToothGrowth[ToothGrowth$supp == 'VC' & ToothGrowth$dose == 2,]$len
```

### Test results

Now we compare the two supplement types as the dose increases from 0 to 0.5 and from 0.5 to 1 ml.:

```
t.test(dose05_suppOJ, dose05_suppVC, paired=FALSE, conf.level = 0.95)$p.value
```

```
## [1] 0.006359
```

```
t.test(dose1_suppOJ, dose1_suppVC, paired=FALSE, conf.level = 0.95)$p.value
```

```
## [1] 0.001038
```

In both cases (from 0 to 0.5, from 0.5 to 1) there is a significant difference between supplement types, because the p-values are lower than 0.05 at a confidence level of 0.95.

This means that, if the  $H_0$  hypothesis is that there is no difference between the effect of each supplement on mean tooth length, then with these p-values that hypothesis can be rejected.

Let's compare the dose increase from 1 to 2 ml.:

```
from_1to2 <- t.test(dose2_suppOJ, dose2_suppVC, paired=FALSE, conf.level = 0.95)
from_1to2$p.value
```

```
## [1] 0.9639
```

```
from_1to2$conf.int
```

```
## [1] -3.798 3.638
## attr(,"conf.level")
## [1] 0.95
```

The p-value is now greater than 0.05. Also, The confidence interval for mean length increase contains the value 0. We conclude from these two points that we cannot reject the  $H_0$  hypothesis.

It would seem that, for lower doses, the OJ method provides better results in tooth length than the VC method. But, past a certain dose, it doesn't matter how the vitamin is administered.