

Tooth Growth basic exploratory data analysis

Tomasz Jaskula

26 avril 2016

Summary

In this project we are going to analyze the tooth growth data. We will use some confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Analysis

Load libraries

```
library(ggplot2)
```

Load the data

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

Exploring the data

Let's explore data size

```
dim(ToothGrowth)
```

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

Structure 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 ...
```

The column `supp` is a factor containing two levels:

- OJ
- VC

Are there any missing data?

```
sum(is.na(ToothGrowth))
```

```
## [1] 0
```

There is no missing values.

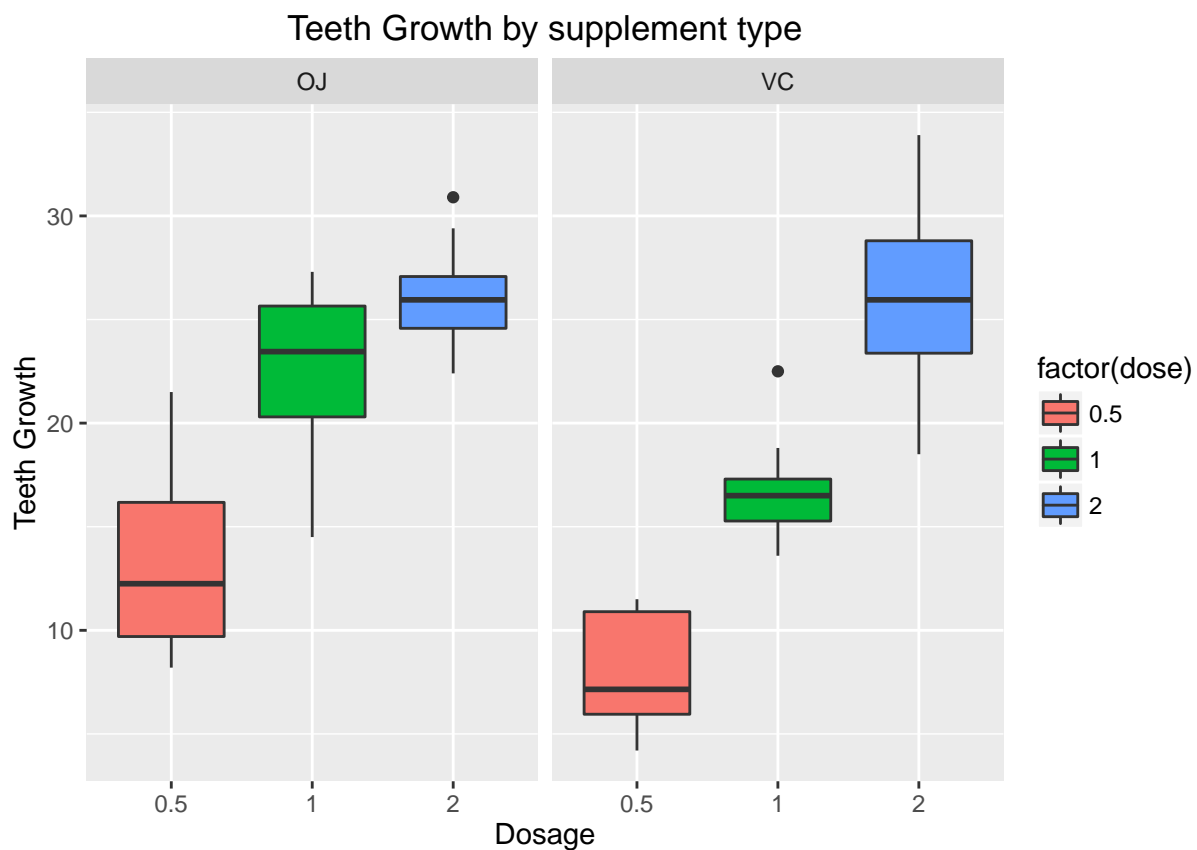
How the head of the data looks like:

```
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
```

Let's see the teeth growth by the dosage

```
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
  geom_boxplot() +
  facet_grid(.~supp) +
  xlab("Dosage") +
  ylab("Teeth Growth") +
  ggtitle("Teeth Growth by supplement type")
```



This analysis show us that the dosage affects the tooth length, the larger the dosage, the longer the tooth.

Confidence intervals and hypothesis testing

This section tries to answer the question if there is a correlation between the tooth growth and the supplement dosage.

The null hypothesis states that there is no correlation between delivery method and tooth length

Dosage as factor We split the data set into 3 datasets, one for each of the doses. This will allow to run the hypothesis test on all the three subsets.

```
dose_05 <-subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose_1 <-subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose_2 <-subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

Let's run the test on each subset

```
t.test(len~dose, data=dose_05, paired=FALSE)
```

```
##
## 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
```

```
t.test(len~dose, data=dose_1, paired=FALSE)
```

```
##
## 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
```

```
t.test(len~dose, data=dose_2, paired=FALSE)
```

```
##
## 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
```

As we see those confidence intervals allows to reject the null hypothesis because there is a correlation between the tooth length and the dosage.

Supplement as factor Let's test the effect of supplement on teeth growth

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)
```

```
##
## 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 doesn't allow to reject the null hypothesis. The supplement type is not correlated to the tooth growth.

Conclusion

Increasing the dose has an effect on tooth growth however the supplement type has not. We assume that the populations are independent.