

# Tooth growth statistical inference

*Stanislav Prikhodko*

*Friday, May 22, 2015*

## Loading data

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(ggplot2)
data(ToothGrowth)
```

## Basic summary of the data

Here is a fragment of data that this study is built on:

```
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
```

Here is some summary of this 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
```

In this study I will analyze the variables *supp* and *dose* that have following values:

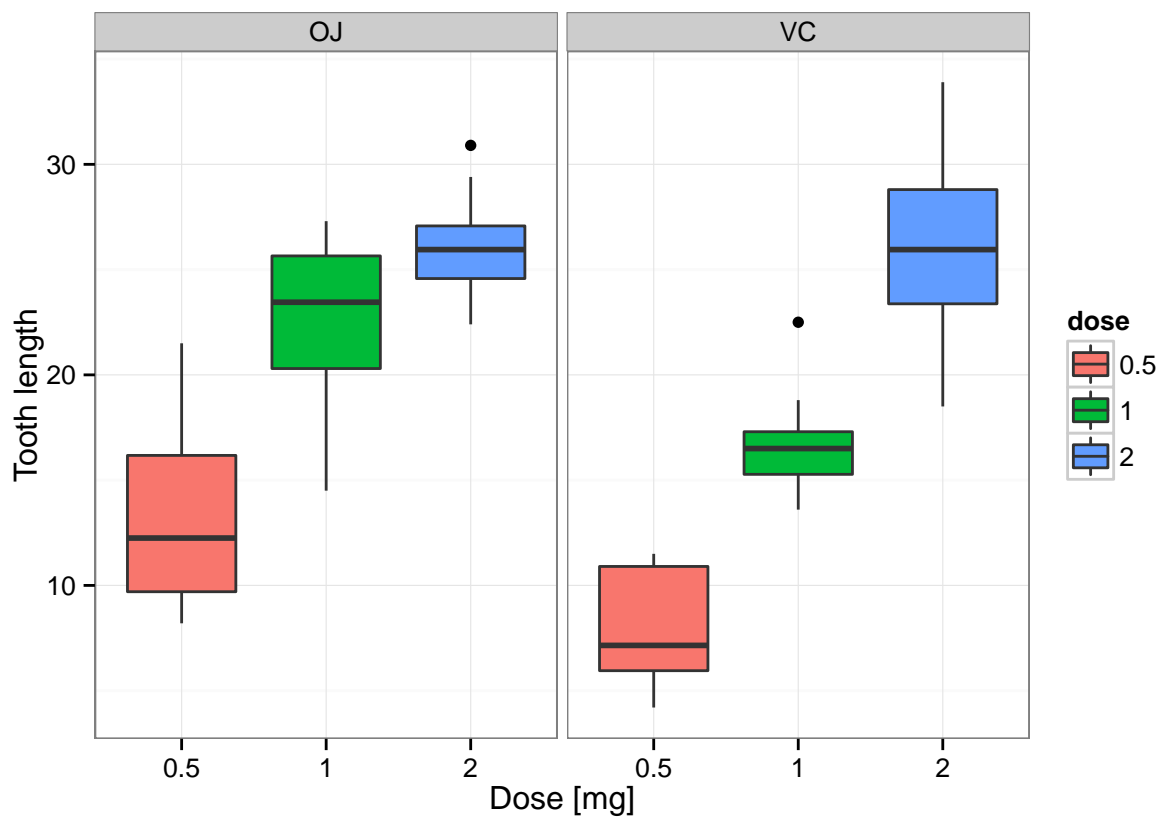
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

## Exploratory data analysis

Looking at values of tooth length for different factors we can notice clear trends.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=dose)) +
  geom_boxplot() +
  theme_bw() +
  facet_wrap(~supp) +
  scale_x_discrete("Dose [mg]") +
  scale_y_continuous("Tooth length")
```



## Assumptions

Since we have that low number of cases in the sample I will use **T-distribution** for the hypothesis testing.

## Confidence intervals and Hypothesis testing

Now, let's use confidence intervals to compare tooth growth by supp and dose. For this research I will prepare subsets of data by different doses:

```
ToothDose1 <- subset(ToothGrowth, dose == 0.5)
ToothDose2 <- subset(ToothGrowth, dose == 1.0)
ToothDose3 <- subset(ToothGrowth, dose == 2.0)
```

Below is t-test for the variable **supp** within all doses and in each of subsets by different doses:

```
t.test(len ~ supp, 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
```

```
t.test(len ~ supp, data = ToothDose1) ## 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
```

```
t.test(len ~ supp, data = ToothDose2) ## dose = 1.0
```

```
##
## 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
```

```
t.test(len ~ supp, data = ToothDose3) ## dose = 2.0
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.0461, 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
```

## Conclusions

Looking at all groups together and the group of samples with dose=2.0 we see that p-value is greater than 0.05 (and the confidence interval contains 0) so we cannot reject the null hypothesis.

However, in the group of samples where dose = 0.5 and 1.0 we are 95% confident that there is a difference between true mean of *tooth length* in groups OJ and VC.