

Statistical Inference Course Project: Part 2

A. Rodríguez

21/9/2020

Overview

On this second part of the course assignment, we're going to analyze the ToothGrowth data in the R datasets package. The steps to follow will be:

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

Analysis and basic summary

First, we'll have to load the dataset from the R package and check its main features:

```
library(datasets)
data("ToothGrowth") ### Load dataset from R package datasets
dim(ToothGrowth) ### It's a dataframe of 60 rows and 3 columns
```

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

```
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
```

```
str(ToothGrowth)
```

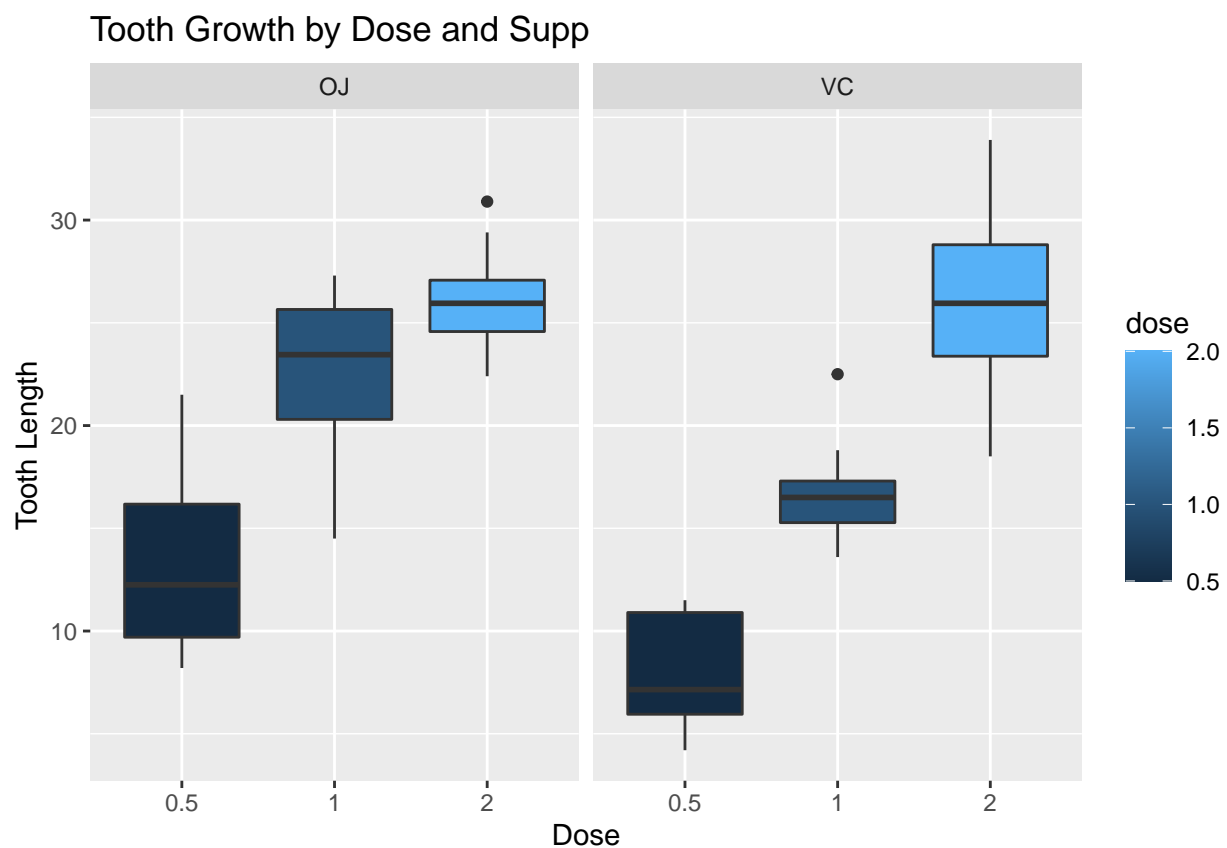
```
## 'data.frame':   60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Basically, this dataset measures different tooth length and divides them by supp (OJ and VC). Then of each supp these measures are classified by dose (0.5, 1.0 and 2.0). The most sensible way to perform exploratory data analysis would be a boxplot.

```
library(ggplot2)
g <- ggplot(data = ToothGrowth, aes(dose, len))
g <- g + geom_boxplot(aes(factor(dose), fill = dose)) + facet_grid(.~supp)
g <- g + labs(title = "Tooth Growth by Dose and Supp", x = "Dose", y = "Tooth Length")
g
```



Confidence tests

When looking at the values of tooth length on the boxplot by supp and disregarding dose, it seems that they spread through a relatively equal range.

```
t.test(data = ToothGrowth, len~supp)[c(3,4)]
```

```
## $p.value
## [1] 0.06063451
```

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

This t-test confirms our suspects. As the p-value > 0.05 and 0 is located within the 95% confidence interval, we can't see any significant difference in tooth length by supp. We can check by pairs of doses (disregarding supp):

```
T1 <- ToothGrowth[ToothGrowth$dose == 0.5,]
T2 <- ToothGrowth[ToothGrowth$dose == 1.0,]
T3 <- ToothGrowth[ToothGrowth$dose == 2.0,]
```

```
T1_T2 <- rbind(T1,T2)
t.test(data = T1_T2, len~dose)[c(3,4)]
```

```
## $p.value
## [1] 1.268301e-07
##
## $conf.int
## [1] -11.983781 -6.276219
## attr("conf.level")
## [1] 0.95
```

```
T1_T3 <- rbind(T1,T3)
t.test(data = T1_T3, len~dose)[c(3,4)]
```

```
## $p.value
## [1] 4.397525e-14
##
## $conf.int
## [1] -18.15617 -12.83383
## attr("conf.level")
## [1] 0.95
```

```
T2_T3 <- rbind(T2,T3)
t.test(data = T2_T3, len~dose)[c(3,4)]
```

```
## $p.value
## [1] 1.90643e-05
##
## $conf.int
## [1] -8.996481 -3.733519
## attr("conf.level")
## [1] 0.95
```

In all 3 t-tests, p-value is effectively non-existing, and the 0 value is well outside the 95% confidence value.

Conclusions

To recap all the information of the t-tests performed, the type of supp applied doesn't seem to affect significantly the tooth growth. On the other hand, the dose apparently has a huge impact on the tooth length, seeing significant differences between doses.