Statistical Inference Project 2

Francisco Martín

October 8, 2018

Overview

This is the part 1 of the Statistical Inference final project. It consist on a study of the R dataset "ToothGrowth", which presents measurements of 60 Guinea Pigs, each 10 at each Vitamin C dose level (0.5, 1 and 2 mg) for each type of delivery method (orange juice or ascorbic acid).

Analysis

Summary of the data

First of all data should be loaded. Let's look how is it like:

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

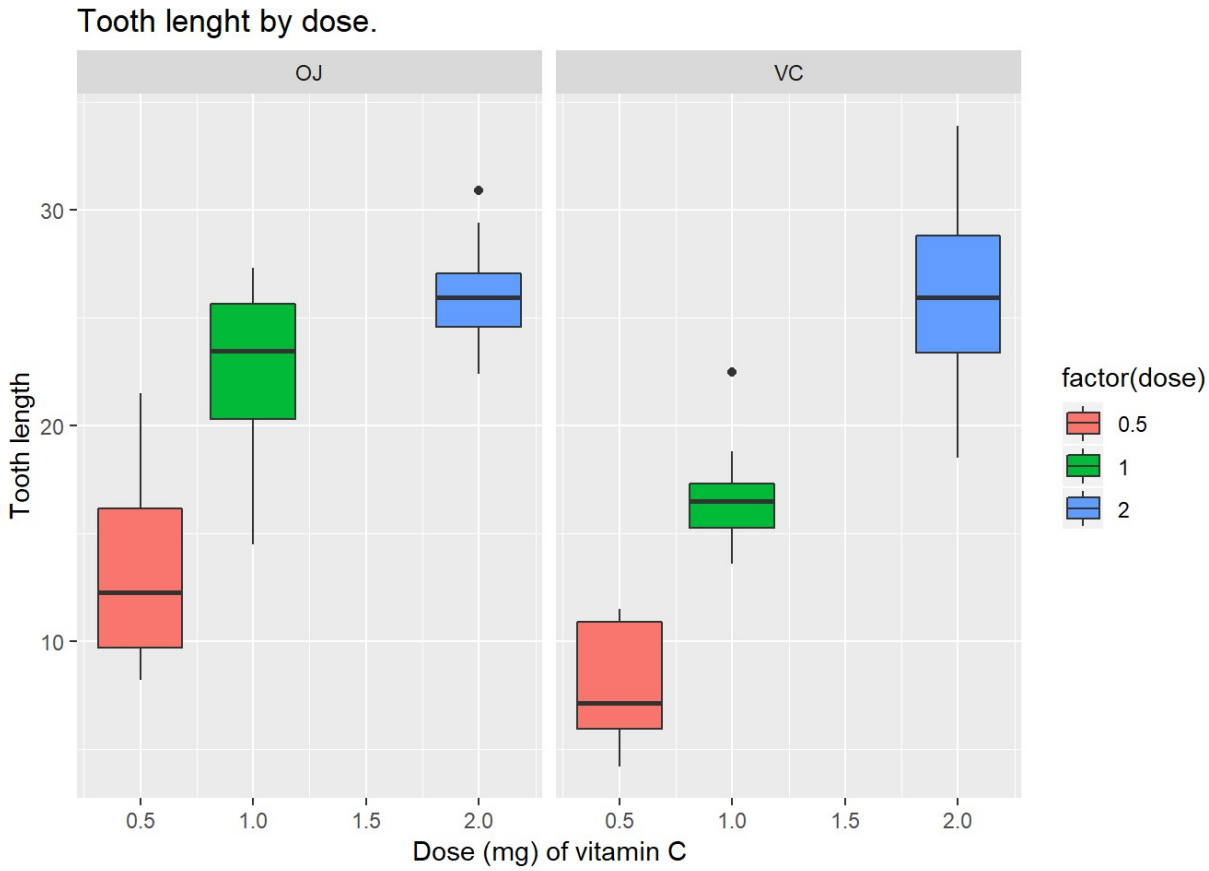
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
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
```

There are 30 observations of every type of supply (OJ orange juice or VC ascorbic acid). as there are 3 different doses, we can use boxplot to see every one of them

```
g <-  ggplot(ToothGrowth, aes(x=dose, y=len))
g + geom_boxplot(aes(fill=factor(dose)))+facet_grid(.~supp)+xlab("Dose (mg) of vita
min C")+ylab("Tooth length")+ggtitle("Tooth lenght by dose.")
```



Hypothesis

Once we have seen how data is distributed, let's do some hypothesis. Hypothesis I think is reasonable here is that orange juice has a significant impact on tooth lenght compared to ascorbic acid. This will be my zero hipothesis H_0. The non-zero hypothesis will be just the opossite: orange juice has no significant impact compared to ascorbic acid. This will be H_a.

Just looking at the data, it looks pretty clear orange juice has a bigger impact when dose is smaller than 2mg. But let's test it using t.test just for be sure of that:

```
t.test(len ~ supp, data=ToothGrowth[ToothGrowth$dose < 2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##           17.965           12.375
```

```
t.test(len ~ supp, data=ToothGrowth[ToothGrowth$dose == 2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, 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 the result of t.test, difference in mean when dose is 2mg is approximately zero, while this difference when dose is lower than 2mg is 5.59. Also, 95% confident interval of the t.test with dose lower than 2 does not include 0, while 95% confident interval of the 2mg test is centered nearly 0. This confirms the hypothesis that orange juice has a bigger impact but only when dose is lower than 2mg.