Tooth Growth, Vitamin C and Orange Juice - A Statistical Approach

Eric Calasans de Barros

December 13, 2017

1 Introduction

In this paper we want study the influence of vitamin C ingest in teeth growth applying some statistical techniques.

2 Metodology

2.1 Loading Data

First we load the ToothGrowt dataset from `dataset` library in R and the others necessary libraries. Here's the code:

```
library(datasets)
library(ggplot2)
library(dplyr)

TG <- data.frame(ToothGrowth)
```

2.2 Exploratory Data Analysis

Next, we can take a look at the data structure, contents and statistical elements such as mean, min, max, etc.

```
str(TG)

## '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 ...

head(TG)

##    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

summary(TG)

##       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
```

We can observe that the dataset is composed by 3 variables:

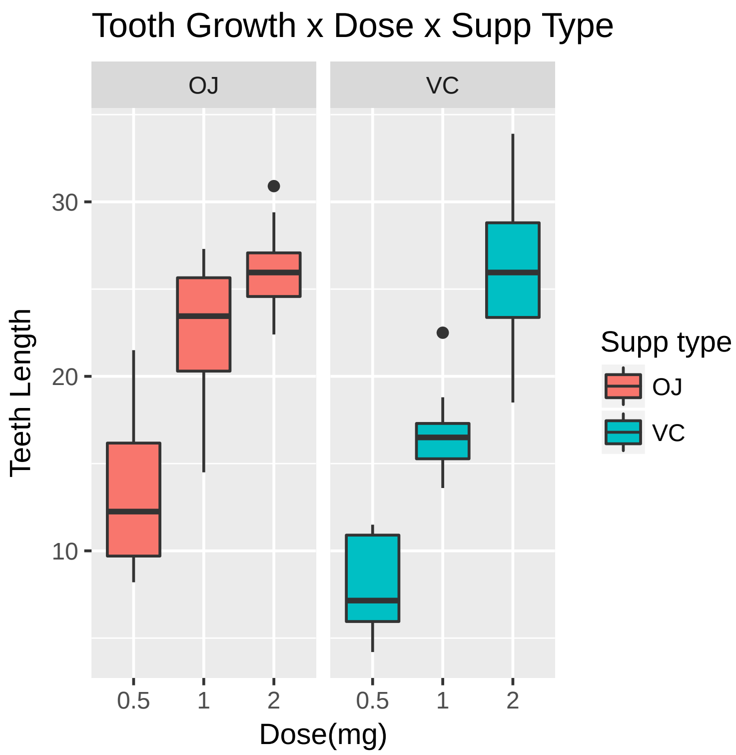
- `len` - tooth length(numerical);
- `supp` - a factor indicating Orange Juice(OJ) or Vitamin C(VC);
- `dose` - dose(mg).

Here we have some exploratory data analysis in the dataset. First, to make easy for plotting, we convert `dose` to a factor variable because there are only three kind of quantity:

```
TG <- mutate(TG, dose = as.factor(dose))
```

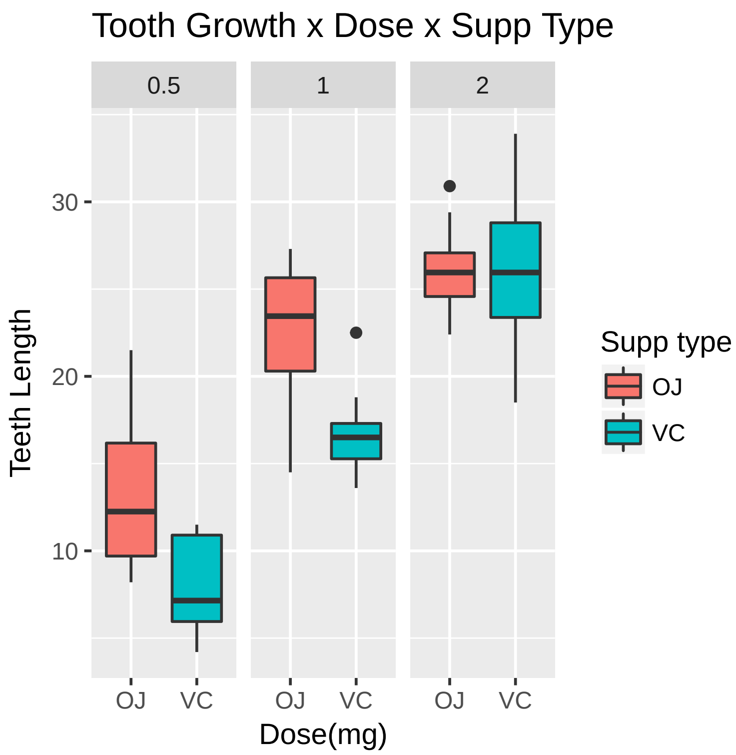
We can make somoe plots and see what we can find:

```
g1 <- ggplot(TG, aes(x=dose, y=len, fill=supp))
g1 <- g1 + geom_boxplot() + facet_grid(.~supp)
g1 <- g1 + ggtitle("Tooth Growth x Dose x Supp Type")
g1 <- g1 + labs(x = "Dose(mg)", y = "Teeth Length")
g1 <- g1 + guides(fill=guide_legend("Supp type"))
g1
```



As far we can observe, when dose increases the tooth length increases too, no matter what kind of supplement is given. Now let's compare supplement type closely.

```
g2 <- ggplot(TG, aes(x=supp, y=len, fill=supp))
g2 <- g2 + geom_boxplot(aes(fill=supp)) + facet_wrap(~ dose)
g2 <- g2 + ggtitle("Tooth Growth x Dose x Supp Type")
g2 <- g2 + labs(x = "Dose(mg)", y = "Teeth Length")
g2 <- g2 + guides(fill=guide_legend("Supp type"))
g2
```



In this plot we can see that low doses of OJ increases tooth growth comparing with VC, which is as effective as OJ in higher doses.

2.3 Hypothesis Test

Hypothesis Tests are a powerful tool in data analysis when we want achieve statistical support for certain statements. For this dataset we want verify influence of kind of supplement and dose amount in tooth growth in a more detailed way. We will test first if kind of supplement influences tooth growth and after that, given one of the doses if one or more is involved in incresase of growth.

For this purpose, given the fact that we don't know population mean and population variance, the **t-test** is indicated.

2.3.1 Testing Supplements

Given $\alpha = 0.05$ here is hypothesis test for supplements:

- H_0: supplement changes tooth growth;
- H_a: supplement doesn't change tooth growth.

```
len <- TG$len
suppl <- TG$supp
t.test(len~suppl, paired=FALSE)

##
##  Welch Two Sample t-test
##
## data:  len by suppl
## 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
```

We observe $p = 0.06063 > \alpha$ and we can reject H_0 given α.

2.4 Testing Doses

For this hypothesis test we need to test each dose(0.5, 1 and 2) separately and check the outcomes. Here is the statments:

- H_0: increasing doses doesn't change tooth growth;
- H_a: increasing doses changes tooth growth.

And the tests:

- **Dose = 0.5 and 1**

```
len1 <- filter(TG, dose == c(0.5, 1)) %>% select(len)
dose1 <- filter(TG, dose == c(0.5, 1)) %>% select(dose)
t.test(len1$len~dose1$dose, paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  len1$len by dose1$dose
## t = -4.4725, df = 17.976, p-value = 0.0002952
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.43327  -5.20673
```

```
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.63             20.45
```

- **Dose = 0.5 and 1**

```
len2 <- filter(TG, dose == c(0.5, 2)) %>% select(len)
dose2 <- filter(TG, dose == c(0.5, 2)) %>% select(dose)
t.test(len2$len~dose2$dose, paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  len2$len by dose2$dose
## t = -7.3335, df = 17.635, p-value = 9.362e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.72833 -10.93167
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.63             25.96
```

- **Dose = 1 and 2**

```
len3 <- filter(TG, dose == c(0.5, 2)) %>% select(len)
dose3 <- filter(TG, dose == c(0.5, 2)) %>% select(dose)
t.test(len3$len~dose3$dose, paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  len3$len by dose3$dose
## t = -7.3335, df = 17.635, p-value = 9.362e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.72833 -10.93167
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.63             25.96
```

We observe that all *p-values* are less than α so we can reject the null hypothesis with $\alpha = 0.05$. By these hypothesis tests we conclude that increasing doses of supplement, no matter what, implies in increase of tooth growth.

3 Conclusion

Statistical Inference tools are very useful, since well-handled, when we intend to change raw data in information.