

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)

## Warning: package 'dplyr' was built under R version 3.4.2

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:

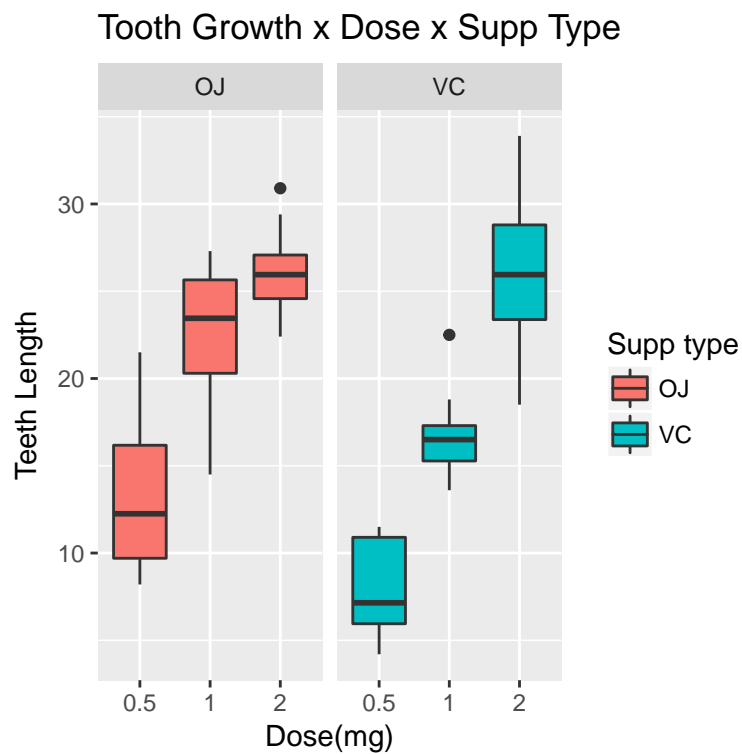
- `len` - tooth length(numerical);
- `supp` - a factor indicating Orange Juice(OJ) or Vitamin C(VC);
- `dose` - dose(mg).

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))
```

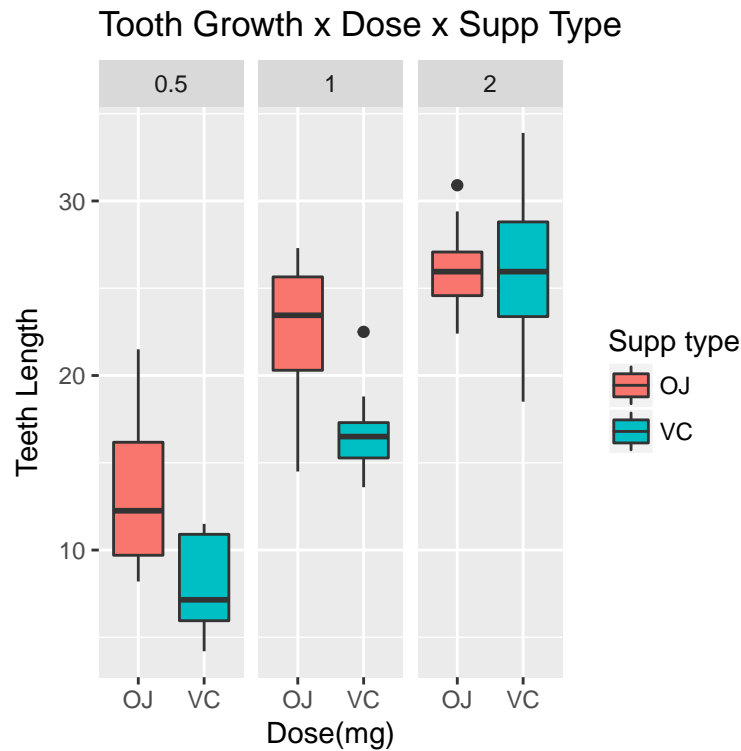
We can make some plots and see what we can find:

```
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 Hypotesis Test

Hypotesis 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.