# Vitamine C increases the tooth growth in ginea pigs

*Juan Sebastián Beleño Díaz*

*31 de enero de 2016*

## Overview

This paper shows important discoveries about the implication of vitamine C in tooth growth in the population of ginea pigs, to do this job was necessary a use of basic concepts in exploratory data analysis and statistic inference.

## Basic exploratory analysis

In this section it'll be shown some metadata related to the TowthGrowth dataset that will be used to perform some analysis in this project, before starting the project is very useful to know about the size of the dataset, its rows and columns, its format, etc.

```
# Load the libraries needed and ToothGrowth data
library(datasets)
library(ggplot2)
data(ToothGrowth)

# Quick look at the dataset variables, the size of the dataset, and the type of
# each variable, and some of the values that each variable contains
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 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

As we can see in this part the ToothGrowth dataset is a data.frame that has 3 variables(`len`, `supp` and `dose`), `dose` and `len` are numeric variables and `supp` is a factor variable, the complete dataset has 60 observations, so it means we won't have problems with RAM while doing the processing part. There is a common patern in the `dose` variable that suggest that maybe there is a few bunch of values that could take, for that reason it'll be converted this variable to factor.

```
# dose variable conversion to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

# Quick look of the dataset to analyze if exist a reduced amount of values for
# dose, which is quite probable
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 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

Here we found the first discovery, just exist 3 values for `dose` (0.5, 1 and 2)

## Summary of the data

According to the official documentation available in https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html . *The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).*

Now that we have some information about the dataset we'll obtain a summary of the dataset to analyze the results and see what we can do with that data.

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

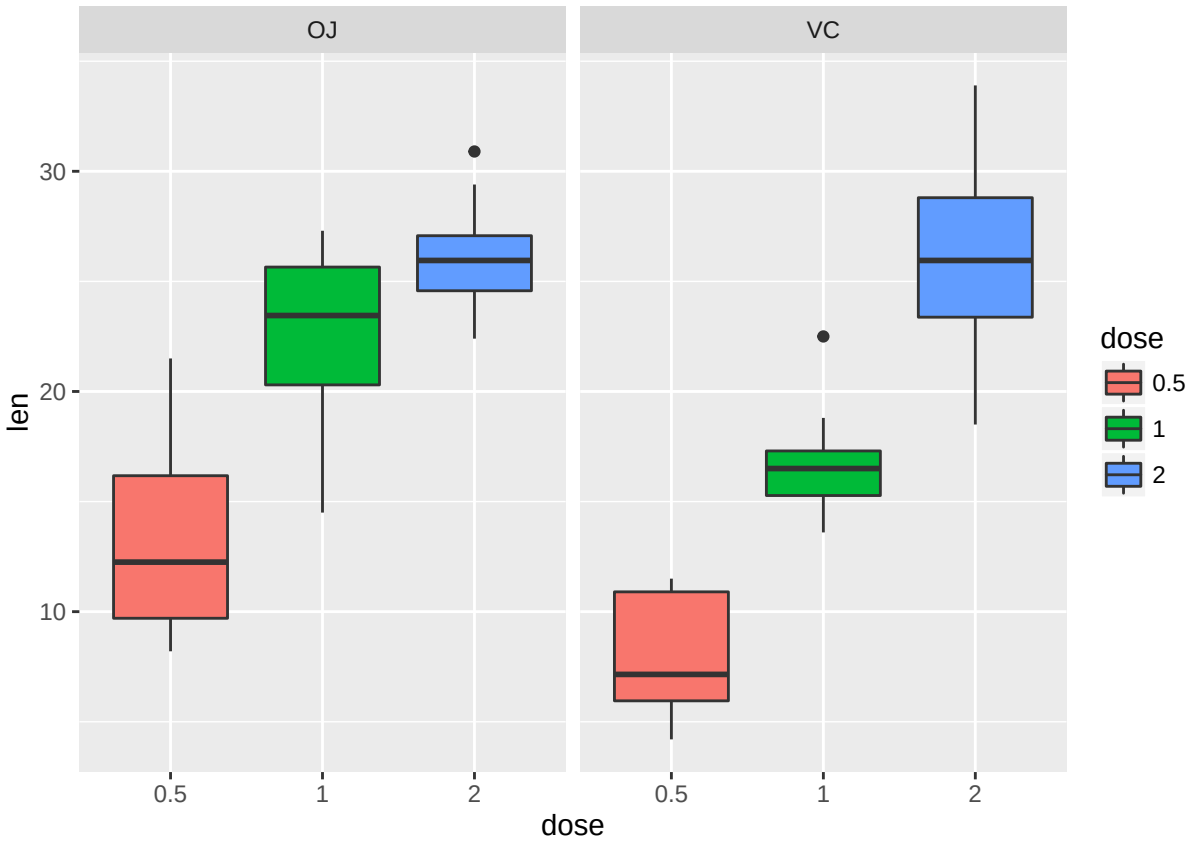
Here we can see that the `len` variable could have values from 4.20 to 33.90 in this sample, the mean is 18.81 and the median 19.25, this data is given for the complete dataset but these `len` values could change taking just one value of `supp` or `dose`.

```
# To see the number of values per combination between dose and supp
table(ToothGrowth$supp, ToothGrowth$dose)
```

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

Indeed we found 10 `len` values for each combination between `supp` and `dose`.

```
g <- ggplot(aes(x = dose, y = len), data = ToothGrowth)
g <- g + geom_boxplot(aes(fill = dose)) + facet_wrap(~ supp)
g
```




In the graphic we can see (without a real proof) that more Vitamine C highly contributes to the tooth growth, and Orange Juice (OJ) seems to be a better presentation in terms of tooth growth than the Ascorbic Acid (VC).

## Hypothesis about tooth growth by supp and dose

The last paragraph in the last section are the hypothesis that will be tested in this part, so first starting by setting $H_0$ and $H_A$.

$H_0$: Vitamine C doesn't contribute to tooth growth

$H_A$: Vitamine C does contribute to tooth growth

To try to prove our hypothesis is necessary to set comparison in Vitamine C levels between 0.5 and 1.0, 1.0 and 2.0 and finally 0.5 and 2.0.

```
# Comparison between 0.5 and 1.0
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(0.5, 1.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
```

```
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
# Comparison between 1.0 and 2.0
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(1.0, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

```
# Comparison between 0.5 and 2.0
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(0.5, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

These results are in favor of $H_A$ with 95% of confidence, the p-values are always under 0.05 and the confidence interval doesn't contain 0.

Having this result let's attack the next hypothesis.

$H_0$: Orange Juice and Ascorbic Acid contribute equal to tooth growth

$H_A$: Orange Juice and Ascorbic Acid doesn't contribute equal to tooth growth

```
# Comparison between Orange Juice and Ascorbic Acid
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
```

This result indicates that we can't reject null hypotesis due to p-values is over 0.05 and the confidence interval contains 0.

### Conclusions

- Vitamine C highly contributes to increase the tooth length in guinea pigs
- There is no difference between give Orange Juice and Ascorbic Acid in the tooth growth in guinea pigs

## Assumptions

- All the ginea pigs were the same age and eat a similar food
- The sample of ginea pigs was random in the general population of ginea pigs
- The sample is representative of genral population if ginea pigs