

# Effect of Vitamin C dose and Delivery Method on Tooth length of Guinea Pigs

*Desiré De Waele*

*10 maart 2016*

## Contents

Synopsis . . . . .	1
Exploratory Analysis . . . . .	1
Hypothesis test for Different Delivery Methods . . . . .	3
Hypothesis test for Different Doses . . . . .	4
Conclusions . . . . .	4

## Synopsis

This assignment explores whether the **delivery methods of Vitamin C and/or the dose has an influence on the tooth lengths of guinea pigs**. The dataset is part of R Base, a documentation of can be found on <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html>. 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)).

This report consists of **three parts**:

1. An exploratory analysis to see if we should even waste our time on hypothesis tests. Boxplots will summarize the data.
2. A hypothesis test testing whether the different delivery methods have an influence on tooth lengths.
3. A hypothesis test testing whether the different Vitamin C doses have an influence on tooth lengths.

**Assumptions** in this report are that the true distribution of tooth lengths of Guinea Pigs are normally distributed. This will result in hypothesis testing with t-values. The **dependencies** for this report are the following packages.

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

## Exploratory Analysis

The first thing we do is loading the data and have a look at it.

```
data(ToothGrowth)
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 ...
```

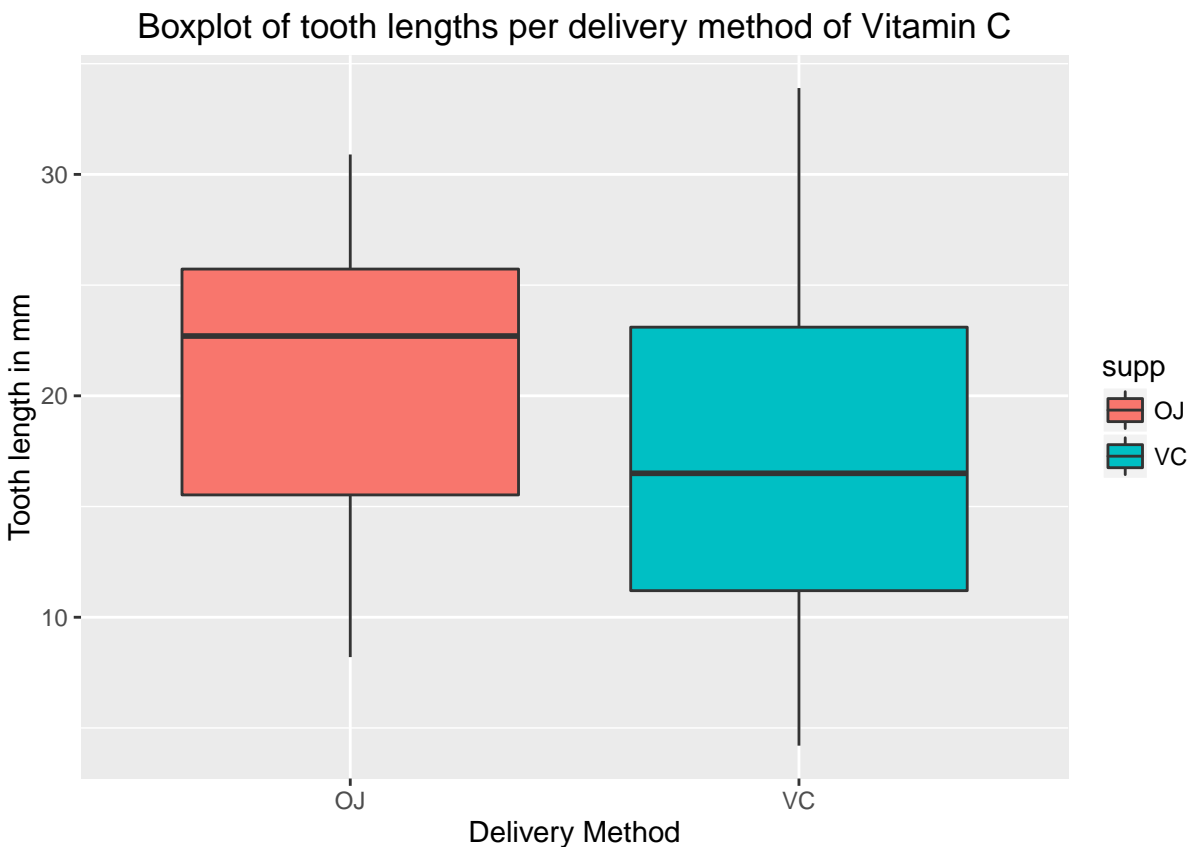
We notice the supplement type is already a factor. Let's make sure the three doses are factors as well. This will presumably help us plot them.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
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
```

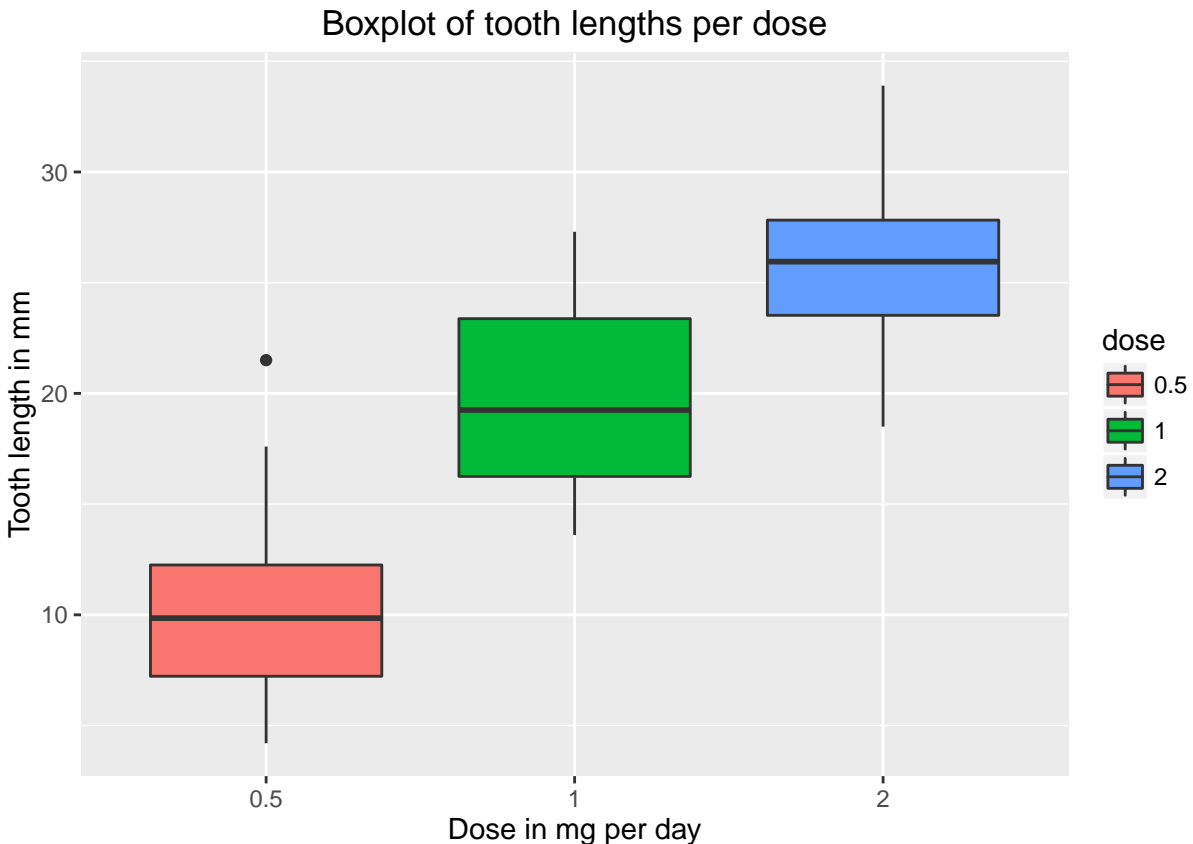
The summary tells us something about the overall data, and how many values there are per factor supplement or dose. We'll now plot the data. First, we group the data per supplement type, i.e. the delivery method.

```
ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot(aes(fill=supp)) +
  labs(title = "Boxplot of tooth lengths per delivery method of Vitamin C",
       x = "Delivery Method", y = "Tooth length in mm")
```



We immediately notice a difference in the median and range of the tooth lengths for both delivery methods. Because of this, it might be worth it to do a hypothesis test for these. Let us check if the same can be said about a possible doses.

```
ggplot(ToothGrowth, aes(dose, len)) + geom_boxplot(aes(fill=dose)) +
  labs(title = "Boxplot of tooth lengths per dose",
       x = "Dose in mg per day", y = "Tooth length in mm")
```



Again, there seems to be a clear correlation, based on the boxplots. This motivates us to do a hypothesis test for the different doses as well. We will do this for the most extreme doses, i.e. a dose of 0.5 mg per day and 2 mg per day.

## Hypothesis test for Different Delivery Methods

We first create a nullhypothesis for the mean difference of both delivery methods. We can reject the nullhypothesis if 0 is not part in the 95% interval for the mean difference. Hence, if 0 *is* part of that interval, it might be the case that there is no mean difference for both delivery methods. In that case, we cannot reject the nullhypothesis.

OJ and VC are our subsets of the data, containing only the values with that particular delivery method. Next, we assume the variance is constant. The pooled variance estimator is stored in `sp`, and is actually just a weighted average both delivery method variances. Next, we plug in `sp` in the interval formula: the difference in mean, plus or minus the relevant t-value, times the standard error.

In the last line, we use the `t.test` function, just to dubbelcheck our manual calculations. Note the false equal variance and paired argument.

```
OJ <- ToothGrowth[ToothGrowth$supp == "OJ",]$len
VC <- ToothGrowth[ToothGrowth$supp == "VC",]$len
n1 <- length(OJ); n2 <- length(VC)
```

```
# pooled variance estimator
sp <- sqrt(((n1-1)*sd(OJ)^2 + (n2-1)*sd(VC)^2) / (n1+n2-2))
# interval formula
mean(OJ)-mean(VC) + c(-1,1) * qt(.975, n1+n2-2) * sp * sqrt(1/n1 + 1/n2)
```

```
## [1] -0.1670064 7.5670064
```

```
# t.test function for dubbelchecking
t.test(OJ, VC, paired = FALSE, var.equal = TRUE)$conf
```

```
## [1] -0.1670064 7.5670064
## attr(,"conf.level")
## [1] 0.95
```

Unfortunately, 0 is part of the interval result. This means that we are not 95% sure that there is no difference between the delivery methods. Therefore, we cannot reject the nullhypothesis.

## Hypothesis test for Different Doses

This part is 100% analogous with the previous part. We subset the data containing only the values with that particular dose. Note we only take in account the values for the two extreme doses, 0.5 and 2 mg/day. We again store the pooled variance estimator and plug it in our interval formula. Again, we dubbelcheck with the t.test function.

```
dose5 <- ToothGrowth[ToothGrowth$dose == .5,]$len
dose2 <- ToothGrowth[ToothGrowth$dose == 2,]$len
n1 <- length(dose5); n2 <- length(dose2)

# pooled variance estimator
sp <- sqrt(((n1-1)*sd(dose5)^2 + (n2-1)*sd(dose2)^2) / (n1+n2-2))
# interval formula
mean(dose5)-mean(dose2) + c(-1,1) * qt(.975, n1+n2-2) * sp * sqrt(1/n1+1/n2)
```

```
## [1] -18.15352 -12.83648
```

```
# t.test function for dubbelchecking
t.test(dose5, dose2, paired = FALSE, var.equal = TRUE)$conf
```

```
## [1] -18.15352 -12.83648
## attr(,"conf.level")
## [1] 0.95
```

Here, 0 is not part of our interval result. Therefore, we can say with 95% certainty that there is a difference between the doses.

## Conclusions

The interval results tell us that:

1. There is at least a 95% certainty that the dose does have an influence on toothlength.
2. There is no 95% certainty that the delivery method has an influence on toothlength.