

# Tooth Growth by Supplement and Dose

*Shuai Wang*

*Saturday, April 25, 2015*

## Overview

In this report, I am going to investigate the relation between tooth growth and supplement, as well as between tooth growth and supplement dose. It turns out that orange juice is better than vitamin C in enhancing tooth growth with statistical significance, when unadjusted for potential confounders. Moreover, more dose in supplement leads to longer tooth, and we also find statistical significance when comparing dose 0.5 to 1, and 1 to 2, still unadjusted for other confounders.

## Data Loading and Exploratory Data Analysis

The data used for this investigation is the ToothGrowth data in R. First let us load the data.

```
data(ToothGrowth)
attach(ToothGrowth)
```

Take a look at the basic summary of the data.

```
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

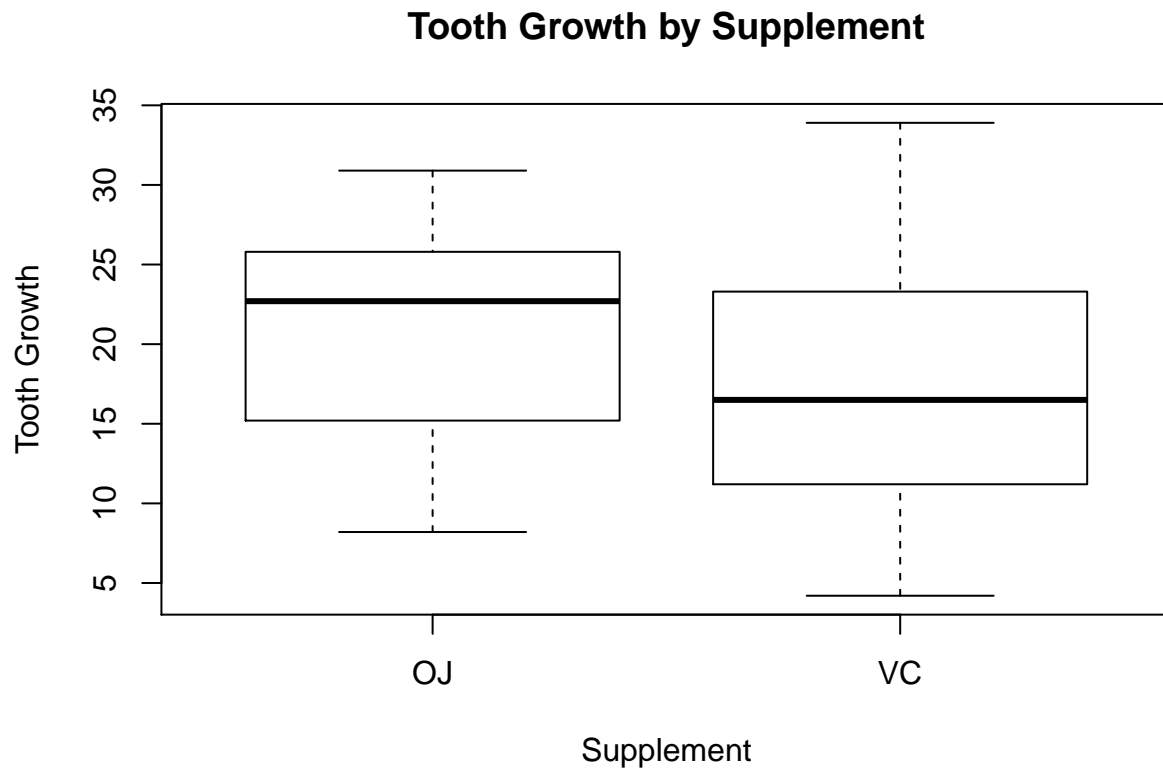
We can see the variable `len` looks like a continuous variable, and I think it indicates the length of tooth growth. The variable `supp` has two values `OJ` and `VC`, and they should be abbreviations of orange juice and vitamin C. Thus `supp` should represent the kind of supplement used. `dose` should be the dose of the corresponding supplement. But from its five number summary (excluding mean), it seems that it is a categorical variable, and we can see that it is true and it has three levels.

```
table(dose)
```

```
## dose
## 0.5  1  2
## 20 20 20
```

Before going to hypothesis tests, I think we'd better explore the data to get an idea of the relations between tooth growth and supplement, as well as between tooth growth and supplement dose.

```
boxplot(len~supp,xlab="Supplement",ylab="Tooth Growth",main="Tooth Growth by Supplement")
```



It looks like that orange juice is better than vitamin C in enhancing tooth growth, and more dose leads to more tooth growth. But those differences may just happen by chance. We need do hypothesis test to confirm them.

## Assumptions and Conclusions

Considering the small sample size, I choose t-test to check the statistical significances between those comparisons. T-test is also highly robust to the underlying assumptions regarding the distribution of the data. It works well whenever the distribution of the data is roughly symmetric and mound shaped, and we can show it is the case for the length of tooth growth for each supplement and supplement dose (see Appendix).

We also need to assume the groups in comparison are independent. I think people taking different supplements and different supplement doses should be different people, thus this assumption is reasonable. Furthermore, we need assume the variances of the length of tooth growths in the comparison groups are unequal, because we don't have any information about it, and we'd better assume unequal variances. Under this assumption, the related standardized statistic is not t-distributed, but it can be approximated by a t-distribution with a fractional degree of freedom.

Now we are going to do hypothesis tests between the groups we are interested. To perform them, we need have null hypotheses. To show significantly statistical differences between groups, I think I'd better choose equal average lengths of tooth growth as the null hypothesis. For alternative hypotheses, because I already see which groups have longer tooth growth and what I am interested now is to know if those differences are statistically significant, thus I choose one-sided hypotheses for them.

First look at the relation between tooth growth and supplement.

```
t.test(len~I(relevel(supp, "OJ")), alternative="greater")
```

```
##
## Welch Two Sample t-test
##
## data: len by I(relevel(supp, "OJ"))
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.4682687 Inf
## sample estimates:
## mean in group OJ mean in group VC
## 20.66333 16.96333
```

So the 95 percent confidence interval of the amount that the average length of tooth growth for taking orange juice exceeds that for taking vitamin C is (0.4682687, Inf), which is above 0. It means that if one had got 30 samples of people taking orange juice and 30 samples of people taking vitamin C repeatedly and randomly from the same population this data is drawn from, about 95 percent of the intervals obtained would be above 0. Thus the average length of tooth growth for taking orange juice is longer than that for taking vitamin C with statistical significance.

It can also be seen from the p-value 0.0303173, which is the probability of obtaining evidence as extreme or more extreme than what we have obtained, under the equal mean hypothesis, i.e., the two averages of tooth growth length for taking orange juice and vitamin C are equal. This probability is very small, and if the null hypothesis is true, then we have observed a rare event. Hence we have much evidence to support that the average length of tooth growth for taking orange juice is longer than that for taking vitamin C.

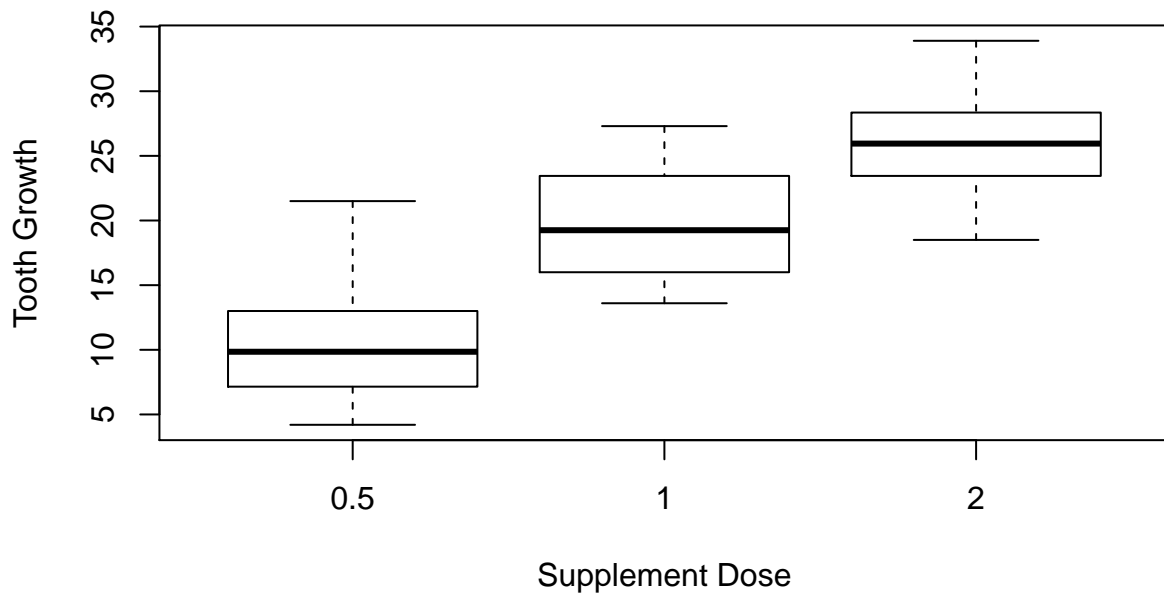
Similarly, we can do hypothesis tests for tooth growth and supplement dose. Because there are three levels of doses, 0.5, 1, and 2, we need to compare them pairwise. Here I only compare dose 0.5 and 1, and leave the other to the Appendix. First let us subset our dataset to the people taking those two doses, and then we do t-test.

```
compare<-ToothGrowth[dose %in% c(0.5,1),]
t.test(len~I(relevel(as.factor(dose), "0.5")), alternative="less", data=compare)
```

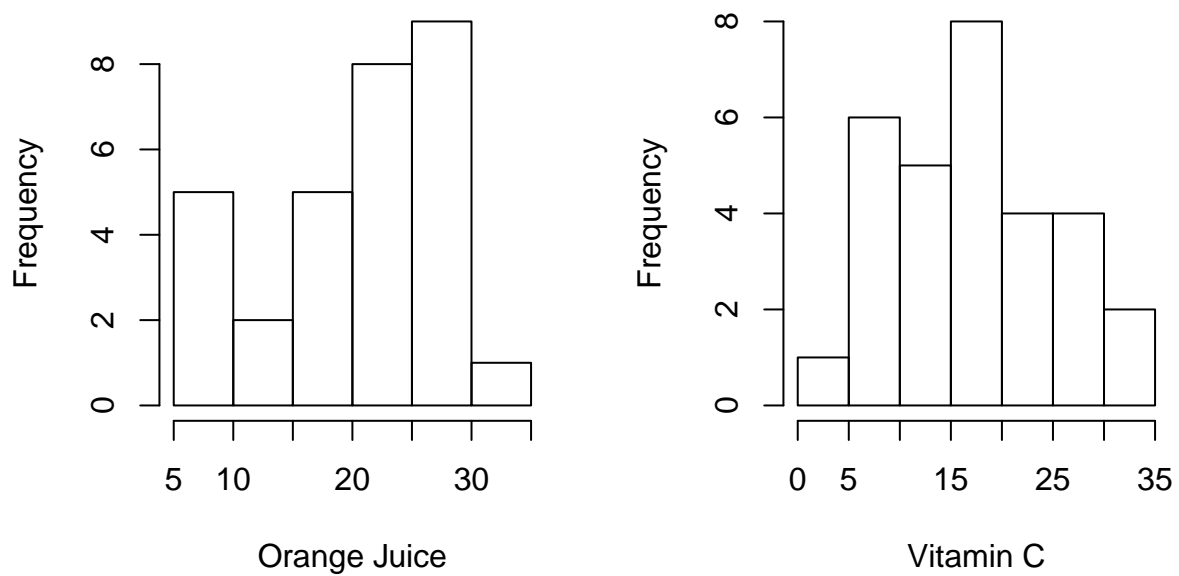
```
##
## Welch Two Sample t-test
##
## data: len by I(relevel(as.factor(dose), "0.5"))
## t = -6.4766, df = 37.986, p-value = 0.00000006342
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
## -Inf -6.753323
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

Same reason as above, the 95 percent confidence interval is (-Inf, -6.7533227), which is below 0, and the p-value is less than 0.0000001, a very small number. Thus if one had got 20 samples each from either dose group repeatedly, about 95 percent of the intervals obtained would be below 0, and the probability of obtaining evidence as extreme or more extreme than obtained under the equal mean hypothesis is less than 0.0000001, a very small number. Either one supports that people taking dose 1 has longer tooth growth in average than those taking dose 0.5.

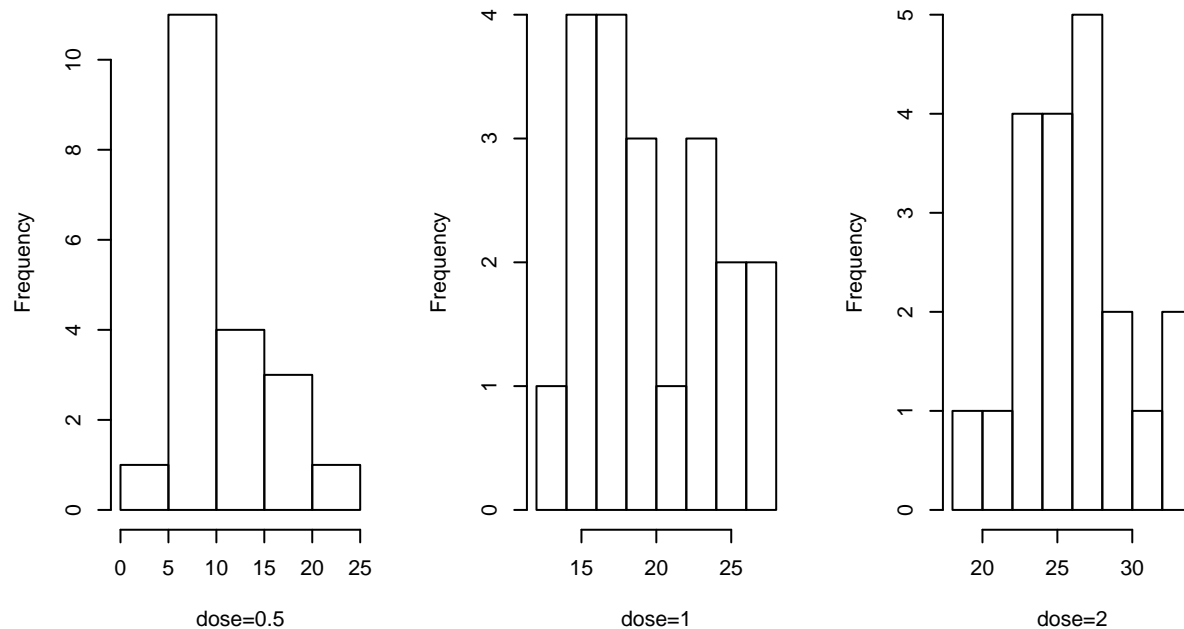
### Appendix A: Tooth Growth by Supplement Dose



### Appendix B: Histogram of Tooth Growth by Supplement



## Appendix C: Histogram of Tooth Growth by Supplement Dose



## Appendix D

It is the t-test performed for tooth growth and supplement dose, when comparing dose 1 and 2. Similar as before, first subset our dataset to the people taking those two doses, and then we do t-test. The t-test output is put here. The 95 percent confidence interval is  $(-\text{Inf}, -4.1738697)$ , which is below 0, and the p-value is less than 0.0000095, a very small number. If one had got 20 samples each from either dose group repeatedly, about 95 percent of the intervals obtained would be below 0, and the probability of obtaining evidence as extreme or more extreme than obtained under the equal mean hypothesis is less than 0.0000095. Either one supports that people taking dose 2 has longer tooth growth in average than those taking dose 1.

```
compare<-ToothGrowth[dose %in% c(1,2),]
t.test(len~I(relevel(as.factor(dose),"1")),alternative="less",data=compare)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by I(relevel(as.factor(dose), "1"))
## t = -4.9005, df = 37.101, p-value = 0.000009532
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean in group 1 mean in group 2
##      19.735      26.100
```