# Investigating The Effect of Vitamin C on Tooth Growth in Guinea Pigs


## Abstract

Vitamin C supplements were given to guinea pigs to determine the effect on tooth growth for different dose sizes (0.5, 1, and 2 mg) and different forms of supplement (orange juice and ascorbic acid). The data show a statistically significant effect for certain pair-wise differences in dosage size and supplement. Higher doses of vitamin C are generally associated with longer tooth length, regardless of the form of supplement. The differences between the forms of supplement is also significant at lower doses, while at higher doses the difference between supplement form is negligible.


## Exploring the Data

The data are provided by the ToothGrowth dataset in the datasets package in R.

```
library(datasets)
data(ToothGrowth)
summary(ToothGrowth)
```

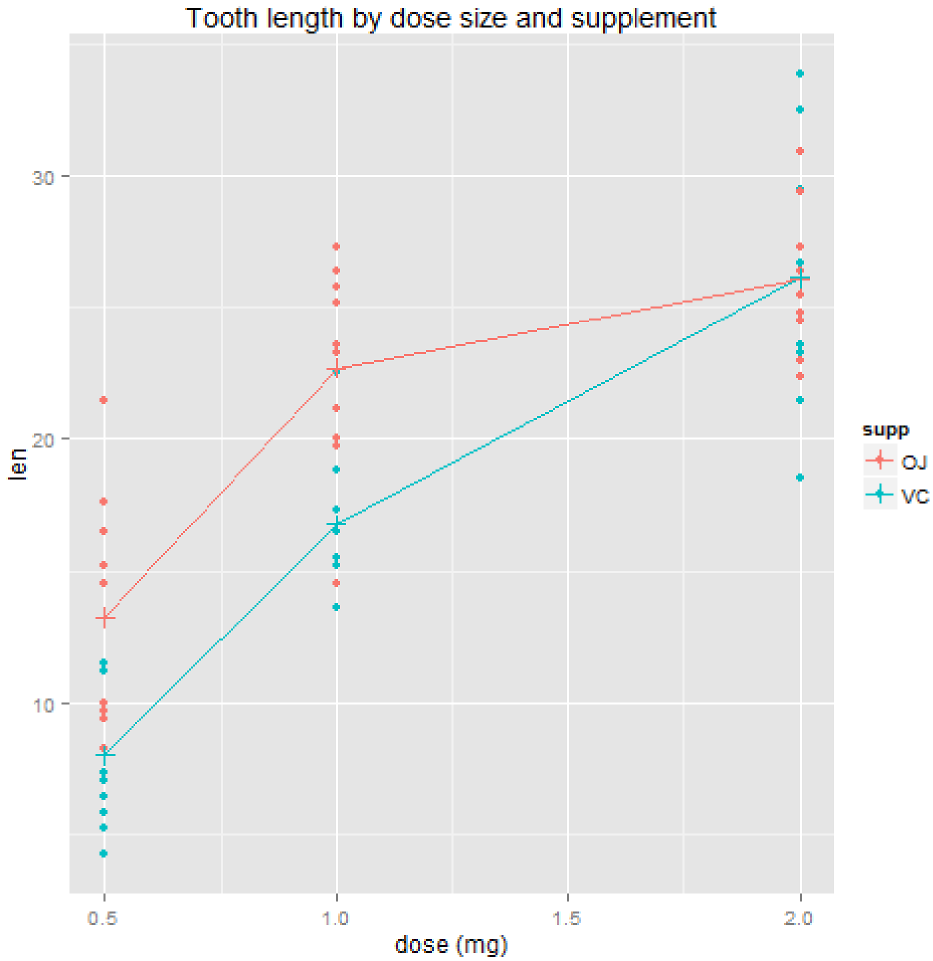
```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

The first chart explores the data by comparing tooth length (len) with dose size (dose). The data are color-coded by supplement (supp), where supp="OJ" is orange juice and supp="VC" is ascorbic acid.

```
library(ggplot2)
library(dplyr)

df <- ToothGrowth %>% group_by(dose,supp) %>% summarize(mean=mean(len))

ggplot(data=ToothGrowth) +
    geom_point(data=ToothGrowth, mapping=aes(x=dose,y=len,colour=supp)) +
    geom_line(data=df,mapping=aes(x=dose,y=mean,colour=supp)) +
    geom_point(data=df,mapping=aes(x=dose,y=mean,colour=supp),shape=3,size=3)
    ggtitle("Tooth length by dose size and supplement") +
    xlab("dose (mg)")
```




I've also drawn cross-hairs connected by lines thru the tooth-length means after grouping by dose and supp. The plot appears to show a significant difference between supplements at low doses, whereas at the largest dose the difference between supplements appears negligible. With respect to dose size, there appears to be a significant difference between 0.5 and 1.0 mg, regardless of supplement. The difference between 1.0 and 2.0 mg appears to be less significant for supp=OJ.

Let's investigate the degree to which the differences between these variables are statistically significant by applying t-tests and calculating confidence intervals .

## Assumptions

Since all guinea pigs were treated and tested independently, we will treat the data groups as independent (unpaired) samples and conduct t-tests between the various groups. There are three variables measured, tooth length (len), dose, and supplement (supp), so for each t-test we must hold one variable constant while testing the other two in order to minimize the effect of confounding variables.

The first question to answer is whether we can treat the variances between the groups as roughly equal when performing the t-test. Let's take a look at within-group variances of each of the 6 groups of data:

```r
# compute variance in each of the 6 groups
aggregate(len ~ dose + supp, ToothGrowth, var)
```

```
##   dose supp       len
## 1  0.5   OJ 19.889000
## 2  1.0   OJ 15.295556
## 3  2.0   OJ  7.049333
## 4  0.5   VC  7.544000
## 5  1.0   VC  6.326778
## 6  2.0   VC 23.018222
```

The variance data shows that while some combinations of variances are roughly equal, others are quite different. So the conservative choice here is to assume the variances between groups are NOT equal.

## Analysis

t-tests are performed between two groups of data, so we must decide between which groups we want to test. Our exploratory analysis above informs us which group differences might be interesting to examine. For one set of tests, we'll compare tooth length difference between supplements while holding the dose size constant. For another set, we'll compare tooth length difference across dose sizes, from 0.5 and 1.0, and from 1.0 and 2.0, while holding the supplement constant.

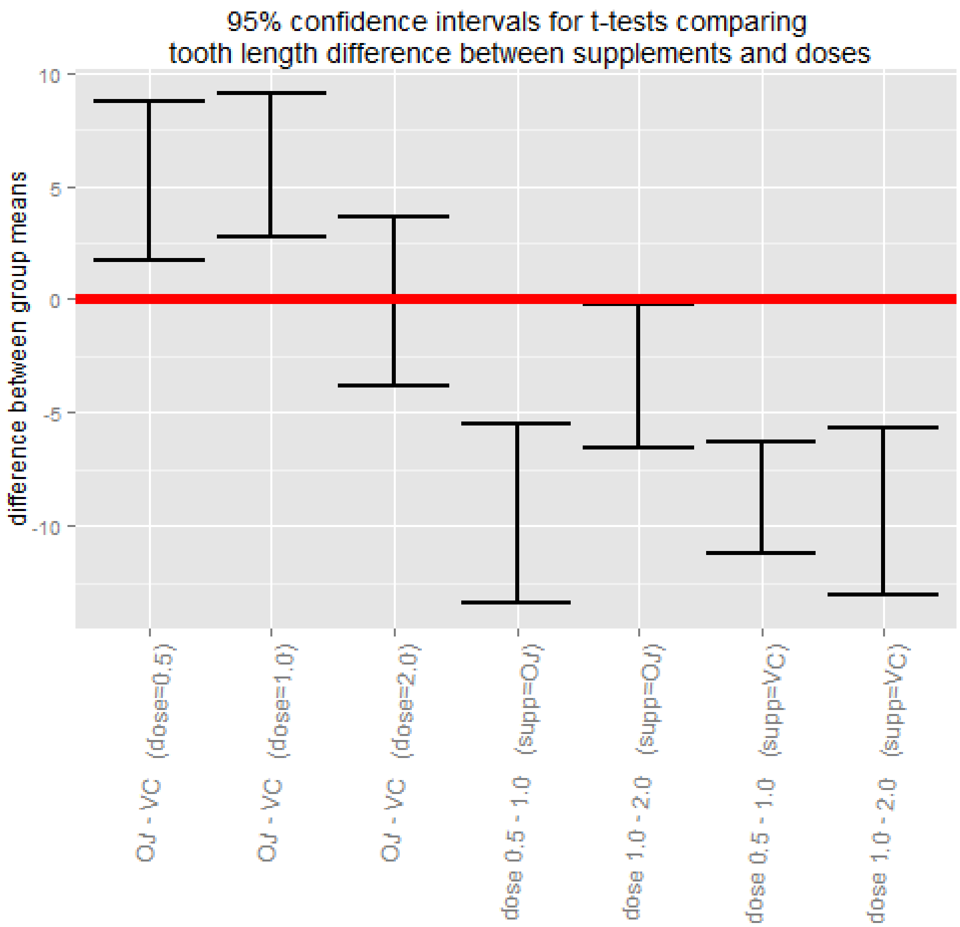
```r
t.tests <- list()
# Compare supplements holding dose constant
t.tests[[1]] <- t.test(len ~ supp,
                       data=subset(ToothGrowth, dose==0.5),
                       paired=F, var.equal=F)
t.tests[[2]] <- t.test(len ~ supp,
                       data=subset(ToothGrowth, dose==1.0),
                       paired=F, var.equal=F)
t.tests[[3]] <- t.test(len ~ supp,
                       data=subset(ToothGrowth, dose==2.0),
                       paired=F, var.equal=F)

# Compare doses holding supplements constant
t.tests[[4]] <- t.test(len ~ dose,
                       data=subset(ToothGrowth, supp=="OJ" & dose %in% c(0.5,
                       paired=F, var.equal=F)
t.tests[[5]] <- t.test(len ~ dose,
                       data=subset(ToothGrowth, supp=="OJ" & dose %in% c(1.0,
                       paired=F, var.equal=F)
t.tests[[6]] <- t.test(len ~ dose,
                       data=subset(ToothGrowth, supp=="VC" & dose %in% c(0.5,
                       paired=F, var.equal=F)
t.tests[[7]] <- t.test(len ~ dose,
                       data=subset(ToothGrowth, supp=="VC" & dose %in% c(1.0,
                       paired=F, var.equal=F)

# extract lower and upper values of the 95% confidence interval
conf.lowers <- sapply(t.tests, function(t) { t$conf.int[1] } )
conf.uppers <- sapply(t.tests, function(t) { t$conf.int[2] } )
conf.mids <- sapply(t.tests, function(t) { (t$conf.int[1] + t$conf.int[2]) /

ggplot() +
    geom_errorbar(mapping=aes(x=factor(1:7),
                              y=conf.mids,
                              ymax=conf.uppers,
                              ymin=conf.lowers),
                  size=1) +
    geom_hline(yintercept=0,color="red",size=2) +
    ggtitle(paste("95% confidence intervals for t-tests comparing\n",
                  "tooth length difference between supplements and doses")) +
    ylab("difference between group means") +
    xlab("") +
    scale_x_discrete(breaks = 1:7, labels=c("OJ - VC   (dose=0.5)",
                                            "OJ - VC   (dose=1.0)",
                                            "OJ - VC   (dose=2.0)",
                                            "dose 0.5 - 1.0   (supp=OJ)",
```

```
                                                         "dose 1.0 - 2.0   (supp=OJ)",
                                                         "dose 0.5 - 1.0   (supp=VC)",
                                                         "dose 1.0 - 2.0   (supp=VC)")) +
    theme(axis.text.x = element_text(angle=85,size=12,hjust=1),
          axis.title.y = element_text(size=12))
```




The red line at y=0 represents zero difference between group means. If the confidence interval includes the red line, it indicates that we cannot reject the null hypothesis; i.e that we cannot say (with 95% confidence) that there is a statistically significant difference between the two group means.

Of the 7 pairs of groups that were tested, 5 pairs show a statistically significant difference between their respective group means, while one pair clearly straddles the red line and another almost does but not quite.

The 5 pairs that show a clear statistically significant difference in group means are:

- orange juice vs. ascorbic acid, holding dose constant at 0.5mg
- orange juice vs. ascorbic acid, holding dose constant at 1.0mg
- dose 0.5 vs 1.0mg, holding supplement=OJ constant
- dose 0.5 vs 1.0mg, holding supplement=VC constant
- dose 1.0 vs 2.0mg, holding supplement=VC constant

The pair that shows no significant difference is orange juice vs. ascorbic acid at dose=2.0mg. We anticipated this from our exploratory analysis which showed virtually no difference between OJ and VC group means at dose=2.0.

The pair with a slight difference is dose 1.0 vs 2.0 while holding supplement=OJ constant. This too we anticipated from our exploratory analysis, which showed a smaller difference going from 1.0 and 2.0 than from 0.5 to 1.0.

## Conclusion

From our results we can conclude with 95% confidence that certain differences between dosage size and supplement have a statistically significant effect on tooth length in guinea pigs.