# Investigating the Effect of Vitamine C on Tooth Growth in Guinea Pigs

Sarah Massengill

2/2/2021

## Overview:

This report goes through a data exploration and performs several hypothesis tests on R's dataset ToothGrowth. In Data exploration we see that as the dose of Vitamin C increases so do the tooth growth lengths. While orange juice outperforms the ascorbic acid at lower doses, the results between the two methods are indistinguishable at 2.0 mg/day. The hypothesis tests using the t-statistic value due to n being on the smaller side, support the hunches made while exploring the data.

## Data Exploration:

According to the R Documentation, ToothGrowth is a data set that detailed the length of a 60 guinea pigs teeth in response to three different doses of either vitamin C given delivered through orange juice (OJ) or an ascorbic acid supplement (VC). The three doses were 0.5,1.0, and 2.0 mg/day. The guinea pigs were grouped first by method, 30 were given OJ and 30 were given VC, and then by dose with 10 guinea pigs in each supplement/dose group.

**Note**: As described in the Help file in R, the tooth length is measured in the "length of odontoblasts" which are cells responsible for tooth growth. Understanding these cells are out of the scope of this project so the lengths will not be labeled with units.

To look at the names of the columns in the data frame a call to the r function names() is made. Below are the names and the R documentation definitions of the meaning for each variable.

- Len = The length of guinea pig tooth growth.
- supp = The supplement given to the guinea pigs: orange juice (OJ), or ascorbic acid (VC).
- dose = The dose given to the guinea pigs: 0.5, 1.0, 2.0 mg/day.

To explore data types, str() is called:

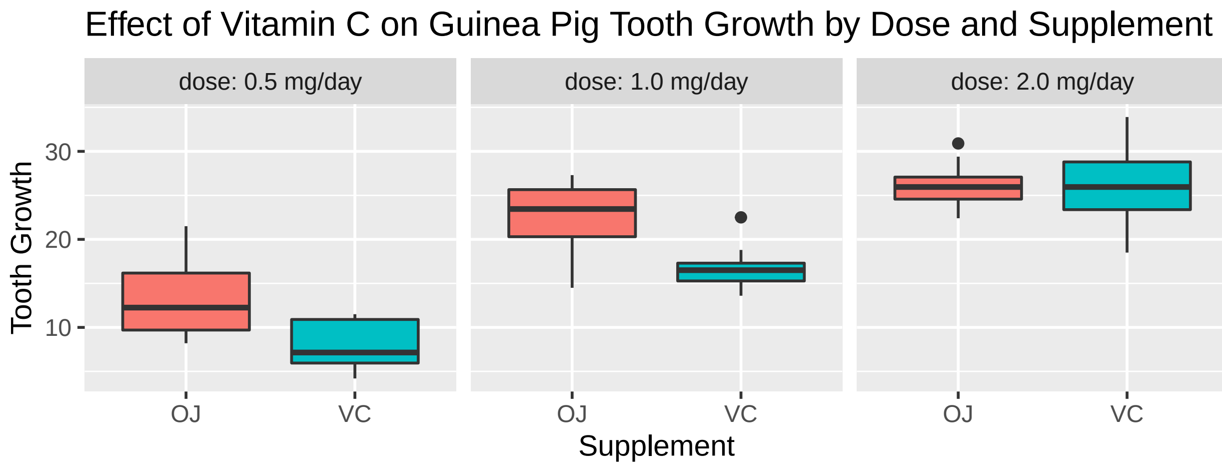
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Here it is clear that there are 60 guinea pigs in the study, the doses and the length are numeric values, and the supplements are Factor values with two levels. Level 1 is OJ and level 2 is VC.

To see how the guinea pigs are split into supplement-dose groups, a call to table is used.

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

There are 10 guinea pigs in each group. Now to visualize the effects the tooth growth for the six groups a call to ggplot's boxplot is applied to the data.




A quick look at the box plots of the data split into dose groups and then into supplement groups leads one to believe that OJ works better on average for the two lower doses, but seems to perform equally well as VC at the dose of 2mg/day. Another thing to notice is that the average increase from 0.5 mg/day to 1.0 mg/day is larger than the average increase from 1.0 mg/day to 2.0 mg/day. This suggests a point of diminishing returns at about 1.0 mg/day, and one might even conclude that if data were collected for a 3mg/day dose that the average increase from 2.0 mg/day to 3.0 may be even less if any. Included in the appendix is a code sample and graph for data visualization from the RStudio documentation. Note the the larger slope between the 0.5 and 1.0 mg/day doses, than the slope between 1.0 and 2.0 mg/day doses.

## Hypothesis Testing

Assumptions for hypothesis tests: * The 60 guinea pigs represent an independent random sample from a normal distribution. * Guinea pigs were randomly assigned to the six dose-supplement groups.

Note: All code and full t.test() outputs can be found in the appendix.

**Test 1**
The null hypothesis is that the dose amount does not matter. Ignoring the supplement given, the doses 0.5 and 1.0 mg/day result in the same average growth. The Alternative Hypothesis is that the larger dose of 1.0 mg/day does in fact effect the average tooth growth.

$$H_o : \mu_{0.5} = \mu_{1.0}, \qquad H_a : \mu_{0.5} \neq \mu_{1.0}$$

```
## [1] "results: p-value = 1.27e-07, confidence interval: [6.28, 11.98]"
```

The results of the non-paired two sample t-test with $n = 20$ and an $\alpha = 0.05$ are in favor of rejecting the null hypothesis. With means $\mu_a = 19.735$ and $\mu_o = 10.606$, the p-value of $1.27 * 10^{-7}$ tells us that it is *very* rare that a difference in the mean values of 9 or more would occur if the null hypothesis were true. The confidence interval of $[6.28, 11.98]$ as expected agrees with the p-value determination of rejecting the null since zero is not in the 95% confidence interval.

**Test 2**
The null hypothesis is that the dose amount does not matter. Ignoring the supplement given, the doses 1.0 and 2.0 mg/day result in the same average growth. The Alternative Hypothesis is that the larger dose of 2.0 mg/day does in fact effect the average tooth growth.

$$H_o : \mu_{1.0} = \mu_{2.0}, \qquad H_a : \mu_{1.0} \neq \mu_{2.0}$$

```
## [1] "results: p-value = 1.81e-05, confidence interval: [3.74, 8.99]"
```

The results of the non-paired two sample t-test with $n = 20$ and an $\alpha = 0.05$ is in favor of rejecting the null hypothesis. With means $\mu_{1.0} = 19.735$ and $\mu_{2.0} = 26.10$, the p-value of $1.81 * 10^{-5}$ tells us that it is rare that a difference of 9 or more in the mean values would occur if the null hypothesis were true. The confidence interval of $[3.74, 8.99]$ as expected agrees with the p-value determination of rejecting the null since zero is not in the 95% confidence interval.

**Note**: These two tests reveal that the dose does in fact have a statistical significance in the results of tooth growth.

**Test 3**
The null hypothesis is that the supplement method does not matter when the dose is 0.5 mg/day, both orange juice and asorbic acid have the same average growth. The Alternative Hypothesis is when the dose is 0.5 mg/day the method of administering Vitamin C does effect the average tooth growth.

$$H_o : \mu_{oj_{0.5}} = \mu_{vc_{0.5}}, \qquad H_a : \mu_{oj_{0.5}} \neq \mu_{vc_{0.5}}$$

```
## [1] "results: p-value = 0.0053, confidence interval: [1.77, 8.73]"
```

The p-value for this hypothesis test is 0.0053 which is smaller than .05, and the 95% confidence interval for the change in means is $[1.77, 8.73] which does not contain zero. So there is enough evidence to reject the null hypothesis. And the we can see that the difference between OJ and VC at 0.5 mg/day is positive meaning that OJ has an average tooth growth larger than that of the ascorbic acid supplement. The mean change for Orange Juice is 13.23 while only 7.98 for ascorbic acid.

**Test 4**
The null hypothesis is that the supplement method does not matter when the dose is 1.0 mg/day, both orange juice and asorbic acid have the same average growth. The Alternative Hypothesis is when the dose is 1.0 mg/day the method of administering Vitamin C does effect the average tooth growth.

$$H_o : \mu_{oj_{1.0}} = \mu_{vc_{1.0}}, \qquad H_a : \mu_{oj_{1.0}} \neq \mu_{vc_{1.0}}$$

```
## [1] "results: p-value = 0.00078, confidence interval: [2.84, 9.02]"
```

The p-value for this hypothesis test is 0.00078 which is smaller than .05, and the 95% confidence interval for the change in means is $[2.84, 9.02] which does not contain zero. So, there is enough evidence to reject the null hypothesis. And the we can see that the difference between OJ and VC at 1.0 mg/day is positive meaning that OJ has an average tooth growth larger than that of the ascorbic acid supplement. The mean change for Orange Juice is 22.70 while only 16.77 for ascorbic acid.

**Test 5**
The null hypothesis is that the supplement method does not matter when the dose is 2.0 mg/day, both orange juice and asorbic acid have the same average growth. The Alternative Hypothesis is when the dose is 2.0 mg/day the method of administering Vitamin C does effect the average tooth growth.

$H_o : \mu_{oj_{2.0}} = \mu_{vc_{2.0}}$,
$H_a : \mu_{oj_{2.0}} \neq \mu_{vc_{2.0}}$

```
## [1] "results: p-value = 0.96, confidence interval: [-3.72, 3.56]"
```

This test actually shows a different result. At a dose of 2.0 mg/day the null hypothesis is <u>not</u> rejected. The p-value of 0.9637 is significantly larger that 0.05 and the 95% confidence interval, [-3.56, 3.72] contains zero almost exactly in the middle of the interval. This means that on average the means for both orange juice and ascorbic acid at doses of 2.0 mg/day do not have statistically significant differences.

# Appendix

Names of columns:

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

Table values of supplement-dose for data Exploration:

```
data.summary <-ToothGrowth %>% group_by(supp,dose) %>%
                summarize(mean = mean(len), median = median(len), min = min(len), max = max(len))
```

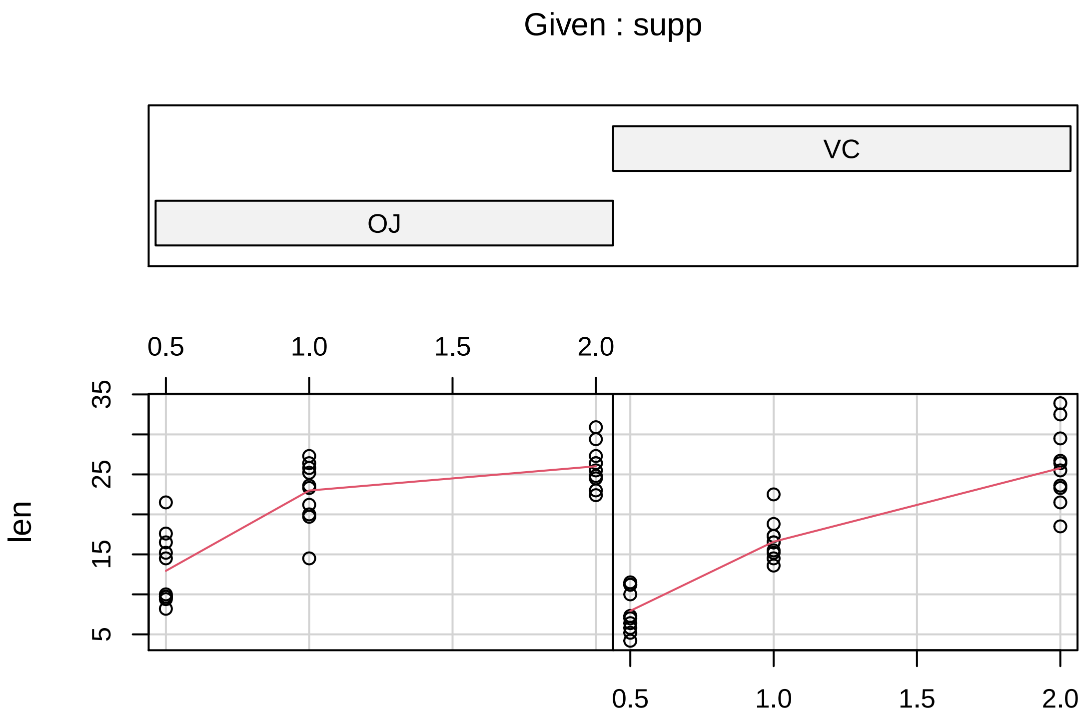
```
## `summarise()` has grouped output by 'supp'. You can override using the `.groups` argument.
```

```
data.summary
```

```
## # A tibble: 6 x 6
## # Groups:   supp [2]
##   supp   dose  mean median   min   max
##   <fct> <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1 OJ      0.5 13.2   12.2    8.2  21.5
## 2 OJ      1   22.7   23.5   14.5  27.3
## 3 OJ      2   26.1   26.0   22.4  30.9
## 4 VC      0.5  7.98   7.15   4.2  11.5
## 5 VC      1   16.8   16.5   13.6  22.5
## 6 VC      2   26.1   26.0   18.5  33.9
```

From RStudio Documentation:

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```




TootbGrowth data: length vs dose, given type of supplement

hypothesis tests:

```
dose0.5<- ToothGrowth %>% filter(dose == 0.5) %>% select(len)
dose1.0<- ToothGrowth %>% filter(dose == 1.0) %>% select(len)
t.test(dose1.0,dose0.5, alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  dose1.0 and dose0.5
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
dose2.0<- ToothGrowth %>% filter(dose == 2.0) %>% select(len)
t.test(dose2.0,dose1.0, alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  dose2.0 and dose1.0
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```r
suppOJ0.5<- ToothGrowth %>% filter(supp == "OJ" & dose== 0.5) %>% select(len)
suppVC0.5<- ToothGrowth %>% filter(supp == "VC" & dose== 0.5) %>% select(len)
t.test(suppOJ0.5,suppVC0.5, alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  suppOJ0.5 and suppVC0.5
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```r
suppOJ1.0<- ToothGrowth %>% filter(supp == "OJ" & dose== 1.0) %>% select(len)
suppVC1.0<- ToothGrowth %>% filter(supp == "VC" & dose== 1.0) %>% select(len)
t.test(suppOJ1.0,suppVC1.0, alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  suppOJ1.0 and suppVC1.0
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.840692 9.019308
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```r
suppOJ2.0<- ToothGrowth %>% filter(supp == "OJ" & dose== 2.0) %>% select(len)
suppVC2.0<- ToothGrowth %>% filter(supp == "VC" & dose== 2.0) %>% select(len)
t.test(suppOJ2.0,suppVC2.0, alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  suppOJ2.0 and suppVC2.0
## t = -0.046136, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.722999  3.562999
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```