The Effects of Acidic Supplements on Guinea Pig Tooth Growth

Salma Rodriguez

Wednesday, March 18, 2015

Introduction

In this study, we analyze observations for the effects of two supplements on tooth growth in guinea pigs. A collection of 60 guinea pigs is grouped into two delivery methods: orange juice (OJ) and vitamin C (VC), and three doses: 0.5 mg, 1 mg, and 2 mg, yielding a total of 10 guinea pigs per group. We would like to have a thorough understanding of the data collected in this study, answering the question: What are the effects of vitamin C and orange juice on odontoblast growth, across three different dose levels?

Exploratory Analysis

First, we load the data.

```
data(ToothGrowth)
```

Here is a summary of 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  
```

We should split the tooth growth by supplement type (supp), to get an idea about the distribution.

```
library(ggplot2)
g <- ggplot(ToothGrowth, aes(x = len))
g <- g + labs(title = "Tooth Growth Across Supplement", x = "Tooth Len")
g <- g + geom_histogram(alpha = 0.20, binwidth = .9, colour = "black", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ supp)
```

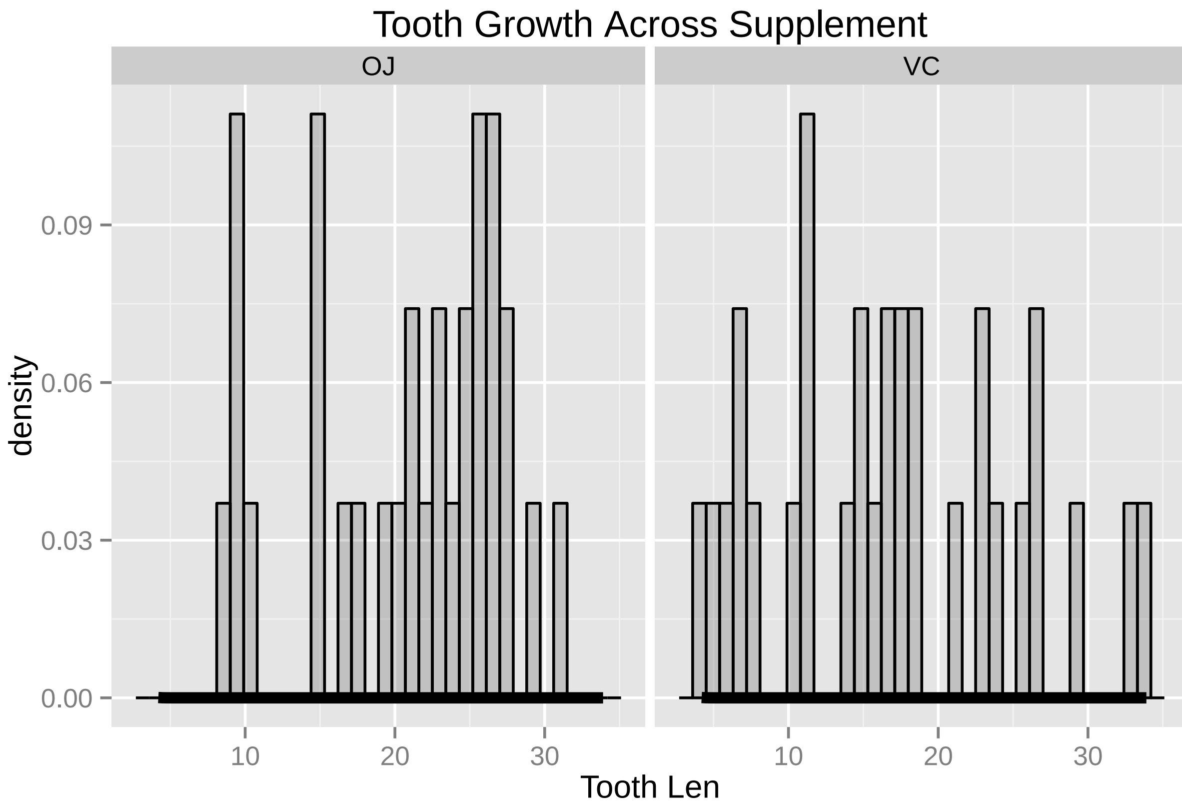

Figure 1: Per Supp

Prima facie, the supplement type (supp) does not seem to have a significant effect on the tooth growth. We will now take the tooth length frequency across dose level, and see if there is any significant difference.

```
g <- ggplot(ToothGrowth, aes(x = len))
g <- g + labs(title = "Tooth Growth Across Dose", x = "Tooth Len")
g <- g + geom_histogram(alpha = 0.20, binwidth = .9, colour = "black", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ dose)
```

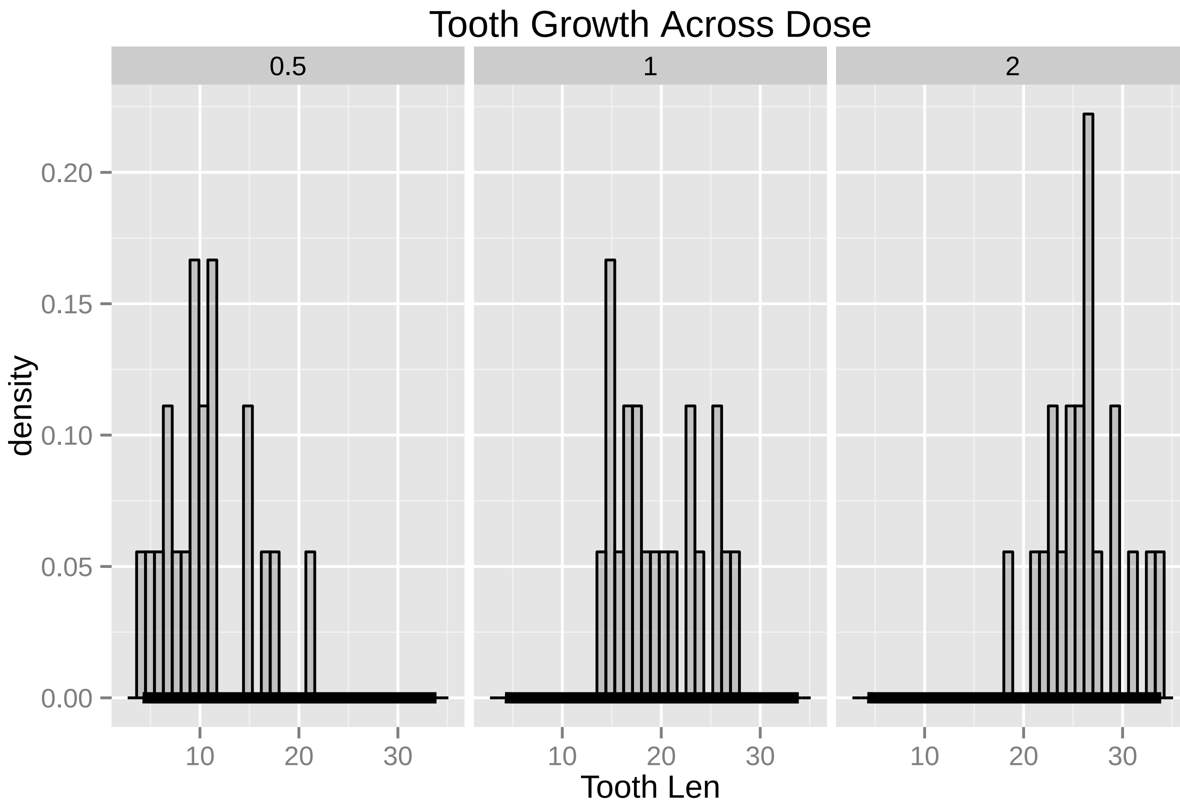

Figure 2: Per Dose

With a higher dose, there is more tooth growth, which is somewhat expected. We would like to see if this correlation is conserved across supplement type and dose level.

```
g <- ggplot(ToothGrowth, aes(x = len))
g <- g + labs(title = "Tooth Growth Across Dose Level & Supplement Type", x = "Tooth Len")
g <- g + geom_histogram(alpha = 0.20, binwidth = .9, colour = "black", aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2)
g + facet_grid(. ~ dose + supp)
```

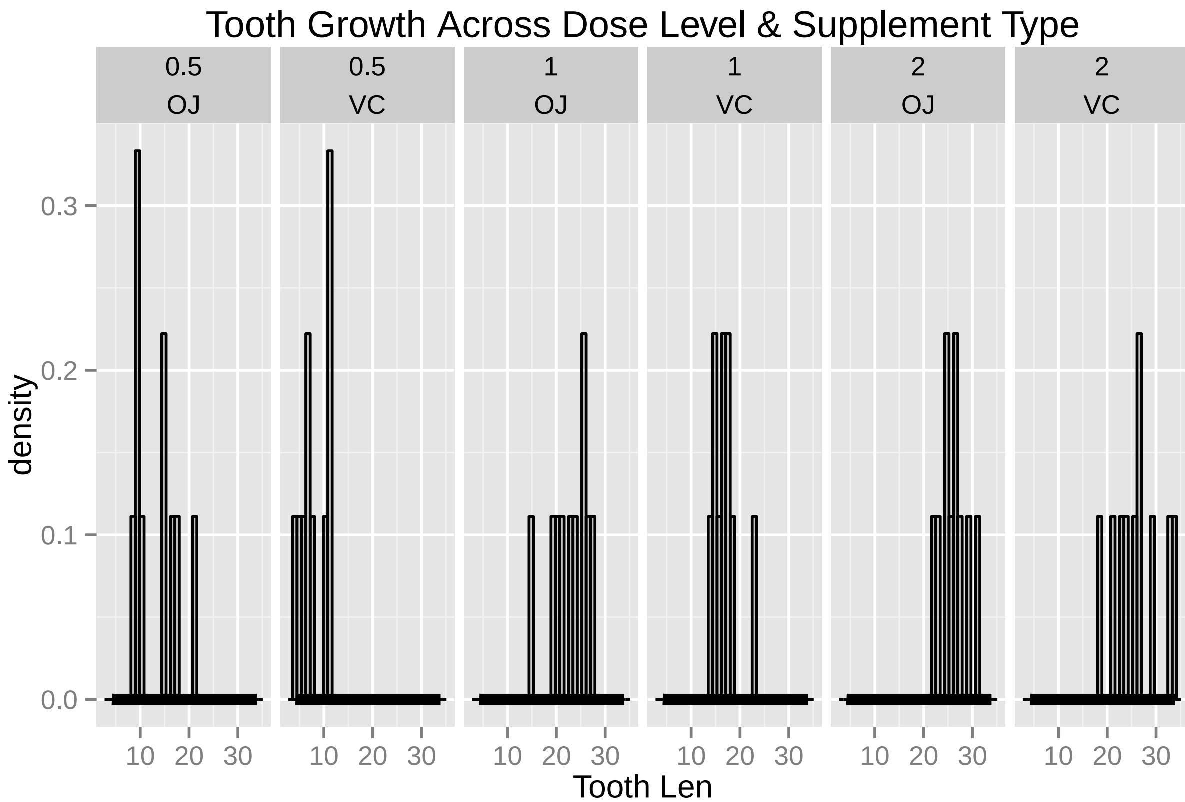

Figure 3: Per Dose & Supp

It seems to be the case that with both OJ and VC there is greater tooth growth with a higher dose. Since **Figure 1** did not reveal anything interesting to us, we shall perform a more thorough investigation, in order to understand which of OJ and VC yield the greatest tooth growth in guinea pigs.

Two-Sample Student's T-test

We will perform a T test for the OJ delivery method versus the VC delivery method, in order to determine how effective orange juice is compared to Vitamin C.

We can separate the OJ supplement and the VC supplement into two groups, having equal sample size.

```
oj = subset(ToothGrowth, supp == "OJ")
vc = subset(ToothGrowth, supp == "VC")
```

After splitting the supplements into separate groups, we can perform further separation, to group by dose level.

```
# for OJ
oj_0.5 = subset(oj, dose == 0.5)
oj_1.0 = subset(oj, dose == 1.0)
oj_2.0 = subset(oj, dose == 2.0)

# for VC
vc_0.5 = subset(vc, dose == 0.5)
```

```
vc_1.0 = subset(vc, dose == 1.0)
vc_2.0 = subset(vc, dose == 2.0)
```

Let us run a t-test, in order to have an overall understanding of what the effect of orange juice is compared to vitamin C. Before we begin the t-test, we ask ourselves two questions about the supplement groups:

1. Can a paired t-test be done?
2. Can a constant variance be assumed?

We note that although the sample sizes are equal, the two groups of guinea pigs are independent groups, since it is not evident that one group relates to the other. Therefore, we will not assume paired sampling.

From **Figure 1** in our exploratory analysis, it is unclear whether the variance is constant for each delivery method. **Figure 2** gives us the frequency of odontoblast growth across dose level, and **Figure 3** gives us the frequency of odontoblast growth across delivery method and dose level. It is evident from these histograms that it would be unsafe to assume equal variances.

Assuming unequal variances and unpaired samples for independent groups OJ and VC, we perform three Welch's t-tests comparing the vitamin C delivery method with the orange juice delivery method across three dose levels: namely, 0.5 mg, 1.0 mg, and 2.0 mg. As a reminder, in order to perform a t-test, we must define a null hypothesis. Our null hypothesis, H_0, will be that the **mean difference** between the OJ supplement and the VC supplement is equal to zero; in other words, that there is no difference in the effect of the two supplements. We will also define an alternative hypothesis, H_a. Both hypotheses can be written as follows:

$$H_0: \quad \mu_{OJ} - \mu_{VC} = 0 \qquad (1) H_a: \quad \mu_{OJ} - \mu_{VC} \neq 0 \qquad (2)$$

Since we assume unequal variances, we can use the Welch-Satterthwaite equation to compute the degrees of freedom:

$$df = \frac{(S_{OJ}^2/n_{OJ} + S_{VC}^2/n_{VC})}{(S_{OJ}^2/n_{OJ})^2/(n_{OJ}-1) + (S_{VC}^2/n_{VC})^2/(n_{VC}-1)} \qquad (3)$$

The Student's t-test requires us to compute a test statistic, t (similar to the z test statistic use with standard normal distributions). Assuming degrees of freedom as equation (3) above, we must first compute the standard error, and mean difference, in order to calculate the test statistic, t.

We can assume the Satterthwaite standard error to be

$$SE = \sqrt{s_{OJ}^2/n_{OJ} + s_{VC}^2/n_{VC}}$$

and the sample mean difference to be

$$XD = \bar{X}_{OJ} - \bar{X}_{VC}$$

We can then compute the t-value as follows:

$$t = \frac{(\bar{X}_{OJ} - \bar{X}_{VC}) - (\mu_{OJ} - \mu_{VC})}{\sqrt{S_{OJ}^2/n_{OJ} + S_{VC}^2/n_{VC}}} \qquad (4)$$

where $\bar{X}_{OJ} - \bar{X}_{VC}$ is the sample mean difference for supplements OJ and VC, $\mu_{OJ} - \mu_{VC}$ is the hypothesized mean difference for supplements OJ and VC. Since the hypothesized mean is assumed to be zero, (4) can be reduced to

$$t = XD/SE = \frac{(\bar{X}_{OJ} - \bar{X}_{VC})}{\sqrt{S_{OJ}^2/n_{OJ} + S_{VC}^2/n_{VC}}}$$

Finally, we may want to compute a 95% confidence interval, if we wish to find out if our hypothesized mean difference is within the confidence interval. This can be done as follows:

$$\bar{X}_{OJ} - \bar{X}_{VC} \pm t_{df,.975} \times SE \qquad (5)$$

We can test the null hypothesis using one of two approaches: checking if the test statistic, $t_{df,.975}$ is between two quantiles: $q_{df,.025}$ and $q_{df,.975}$, or using equation (5) above to figure out if the hypothesized mean difference is covered by the 95% confidence level.

We can manually perform the computation, as above, but luckily, R has a handy t-test function that can be used for performing a Student's t-test. All we would need to test is whether or not the p-value is below 0.025. for the 95% confidence interval.

```
t.test(oj_0.5$len, vc_0.5$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj_0.5$len and vc_0.5$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
t.test(oj_1.0$len, vc_1.0$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj_1.0$len and vc_1.0$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
t.test(oj_2.0$len, vc_2.0$len, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj_2.0$len and vc_2.0$len
## t = -0.0461, df = 14.04, p-value = 0.9639
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The results above clearly indicate that the computed p-value is significant enough to reject the hypothesis that the OJ supplement and VC supplement have equal effect on the tooth growth of guinea pigs, for 0.5 mg and 1.0 mg. For a higher dose of 2 mg, the OJ supplement and VC supplement seem to have equal effect on tooth growth, and therefore we cannot reject the null hypothesis for the 2.0 mg dose level at the 95% confidence level. A careful inspection reveals that with dose levels 0.5 mg and 1.0 mg, it may be the case that orange juice is more effective than vitamin C. However, our t-test can only be used to reject the null hypothesis and cannot be used to draw conclusions about alternatives.

We can confirm the results of our t-test by taking the average tooth growth per dose level in each supplement group. We can do this with a simple function in R.

```r
library(reshape2)
hi <- dcast(ToothGrowth, supp ~ dose, fun.aggregate = mean, value.var = "len")

hi
```

```
##   supp   0.5     1     2
## 1   OJ 13.23 22.70 26.06
## 2   VC  7.98 16.77 26.14
```

This confirms that tooth growth is not equal for the 0.5 mg and 1.0 mg doses. It also seems to indicate that orange juice yields more tooth growth in guinea pigs than abscorbic acid, or vitamin C. However, no conclusion will be drawn without further hypothesis testing.