Impact of Vitaman C Supplements on Guinea Pig Tooth Growth

Chris Keaney

June 19, 2015

Overview

In this analysis we are going to be exploring the impact of Vitamin C supplements on the tooth growth of guinea pigs using the data set **ToothGrowth** from the R datasets package. This data set contains 60 observations of 3 variables:

- len: The measured length of teeth in millimeters.
- supp: The supplement type (VC or OJ)
- dose: The does (in milligrams) provided to the subject (.5 / 1 / 2)

The dataset contains measurements on each of 10 guinea pigs that were given some combination of supplement type and dose, for a total of 60 guinea pigs tested. We will be exploring this data set and running tests on supplement and dose to see if we can determine if there's any significant change in tooth length from the supplement type or dosage.

Data

The following code loads our dataset and converts the column **dose** to a factor variable:

```
require("ggplot2")
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
library(datasets)
teeth <- ToothGrowth

teeth$dose <- as.factor(teeth$dose)

supp_vc <- teeth[teeth$supp == "VC",]
supp_oj <- teeth[teeth$supp == "OJ",]

dose_05 <- teeth[teeth$dose == "0.5",]
dose_1 <- teeth[teeth$dose == 1,]
dose_2 <- teeth[teeth$dose == 2,]
```

The above code creates additional subsets for use later in the analysis.

Exploratory Data Analysis

Now we take a look at our data using a few different methods:

```
summary(teeth)
```

```
##      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
```

As expected, we see we have 60 observations with even splits between supplement type and dosage. Let's now look at the mean, variance, and stadard deviation of the data set:

```
mean(teeth$len)
```
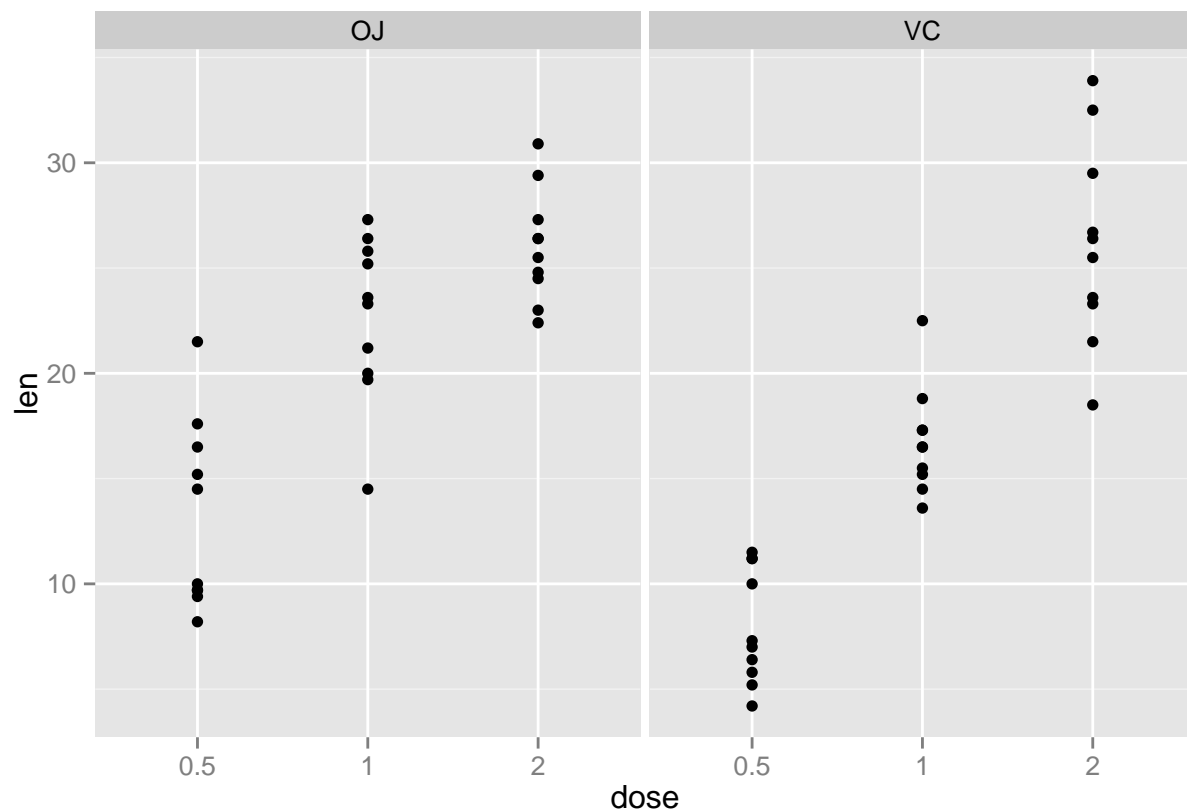
```
## [1] 18.81333
```

```
var(teeth$len)
```

```
## [1] 58.51202
```

```
sd(teeth$len)
```

```
## [1] 7.649315
```

We have a rather large standard deviation on this data set, so let's take a look at a couple of plots using our **supplement** and **dosage** factors.

```
qplot(dose,len,data=teeth,facets=.~supp)   # a quick scatterplot of each delivery method
```

This first plot is a quick scatterplot of our data with a breakdown of both factors. This plot seems to suggest that dose is a significant driver in increased tooth length, and that there may be some significance to the delivery method as well. Let's look at individual box plots of supplement and dose.

```
plot1 <- ggplot(teeth, aes(supp,len))
plot1 <- plot1 + geom_boxplot() + xlab("Supplement") + ylab("Tooth Length (mm)")
print(plot1)
```
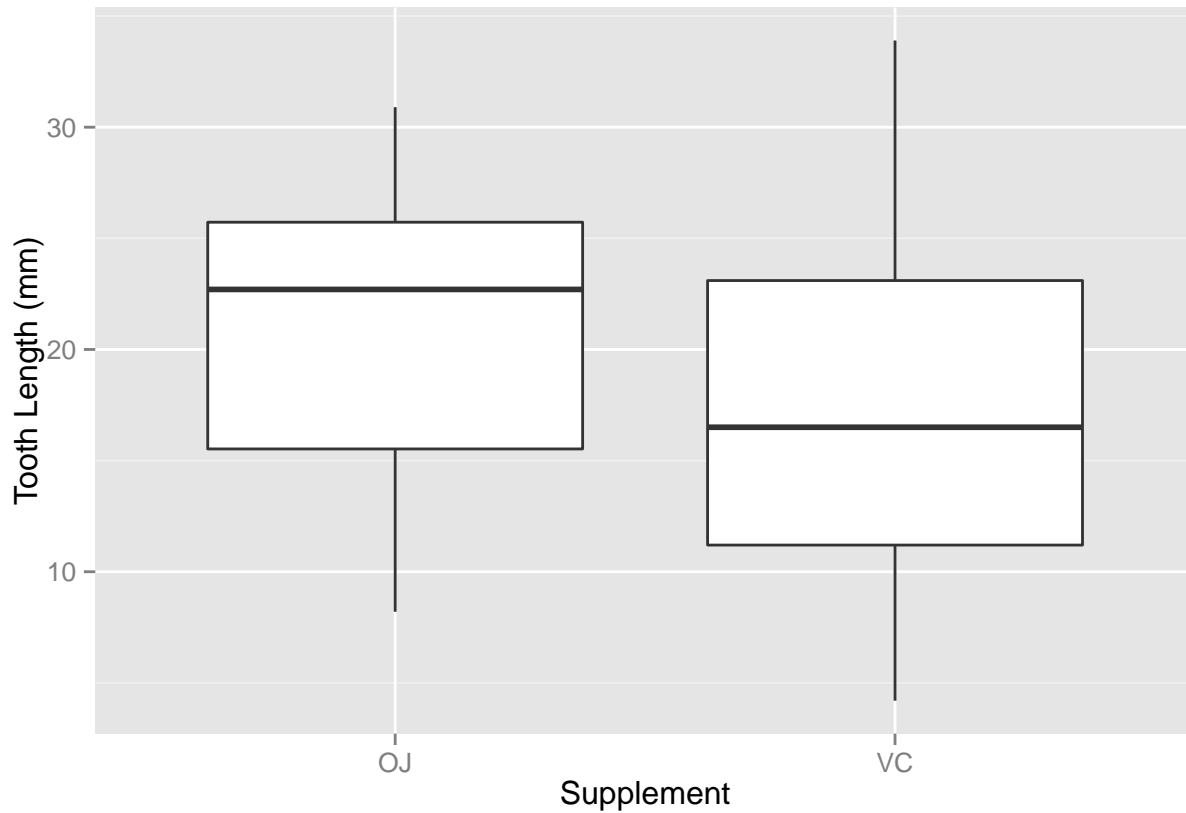


```
plot2 <- ggplot(teeth, aes(dose, len))
plot2 <- plot2 + geom_boxplot() + xlab("Dosage") + ylab("Tooth Length (mm)")
print(plot2)
```
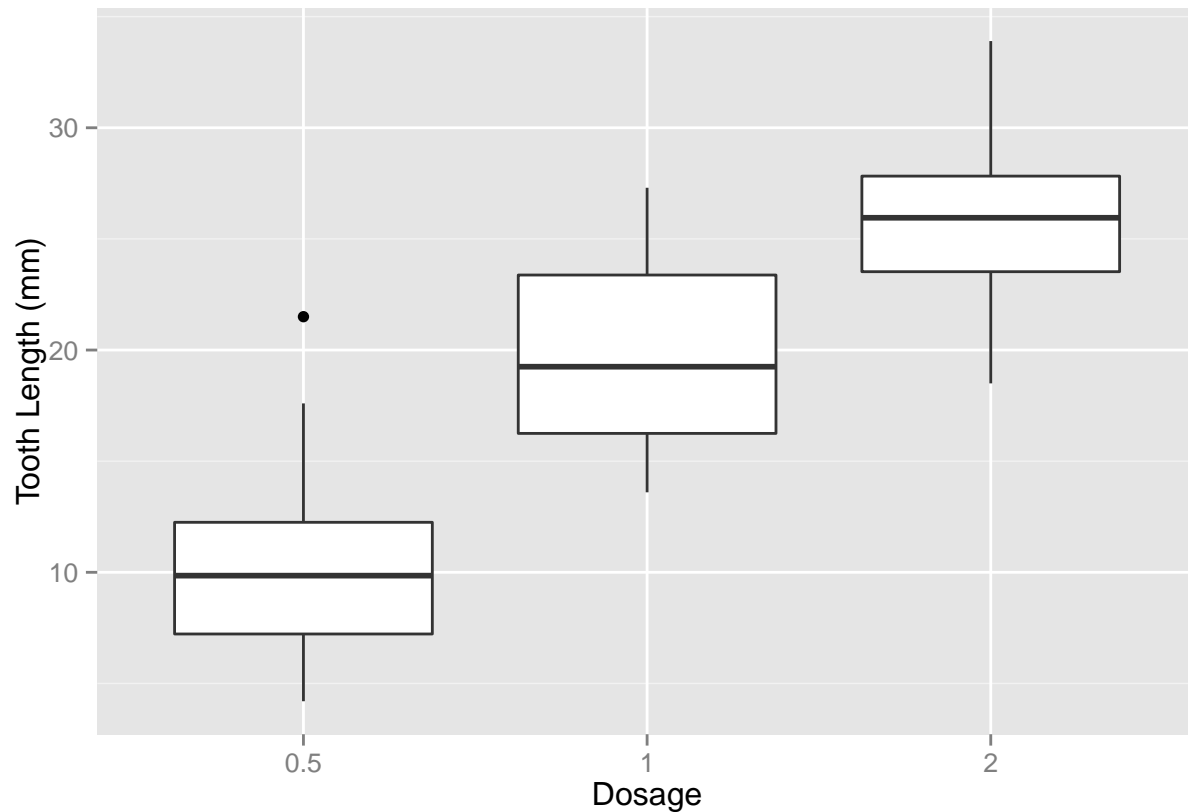
Looking at the above two plots:

1) Supplement appears to have some siginificance in the tooth length measure, but given our sample size, we will test to see if this is statistically significant.

2) The dosage level appears to be a major factor in the tooth length measure, but again we will run some T tests to see if this is a function of the sample or if dosage is in fact a driver.

Hypothosis Tests

We will run a series of T tests against the mean of the overall population to determine if the individual dosage levels or supplement types will allow us to reject the null hypothesis that the mean of the population is approximately 18.9 regardless of supplement or dosage level. We will use in 95% confidence level in all tests.

T Test - Overall Population vs Supplment VC

```
supp_test_1 <- t.test(teeth$len, supp_vc$len, paired=FALSE)$conf.int
supp_test_1
```

```
## [1] -1.765492  5.465492
## attr(,"conf.level")
## [1] 0.95
```

The above confidence includes 0 so we cannot reject the null hypothesis.

T Test - Overall Population vs Supplment OJ

```
supp_test_2 <- t.test(teeth$len, supp_oj$len, paired=FALSE)$conf.int
supp_test_2
```

```
## [1] -4.961865  1.261865
## attr(,"conf.level")
## [1] 0.95
```

The above confidence level includes 0 so we cannot reject the null hypothesis.

T Test - Overall Population vs Dosage 0.5

```
dose_test_1 <- t.test(teeth$len, dose_05$len, paired=FALSE)$conf.int
dose_test_1
```

```
## [1]  5.384549 11.032117
## attr(,"conf.level")
## [1] 0.95
```

The above confidence level is entirely above zero, meaning we can reject the null hypotheses. In this case since we are taking the entire population minus the dosage population, it means that the dosage population produces tooth length results below reasonable expectations given the overall population.

T Test - Overall Population vs Dosage 1

```
dose_test_2 <- t.test(teeth$len, dose_1$len, paired=FALSE)$conf.int
dose_test_2
```

```
## [1] -3.717435  1.874101
## attr(,"conf.level")
## [1] 0.95
```

The above confidence level includes 0 so we cannot reject the null hypothesis.

T Test - Overall Population vs Dosage 2

```
dose_test_3 <- t.test(teeth$len, dose_2$len, paired=FALSE)$conf.int
dose_test_3
```

```
## [1] -9.879846 -4.693487
## attr(,"conf.level")
## [1] 0.95
```

The above confidence level is entirely below zero, meaning we can reject the null hypotheses. In this case since we are taking the entire population minus the dosage population, it means that the dosage population produces tooth length results above reasonable expectations given the overall population.

Conclusions

The tests above show that the dosage level of the supplement has a significant impact in the lenght of teeth in guinea pigs, and that the supplement type does have the same impact. We have made the assumption that our dataset is representative of the overall guinea pig population. While this may not be the case, the results indicate that the higher the dosage of vitamin C (regardless of how it is delivered), the higher the increase in the growth of teeth in guinea pigs.

Appendix

Below is all of the code needed to replicate this analysis:

```r
# load tooth growth data & create subsets for later analysis

require("ggplot2")
library(ggplot2)
library(datasets)
teeth <- ToothGrowth

teeth$dose <- as.factor(teeth$dose)

supp_vc <- teeth[teeth$supp == "VC",]
supp_oj <- teeth[teeth$supp == "OJ",]

dose_05 <- teeth[teeth$dose == "0.5",]
dose_1 <- teeth[teeth$dose == 1,]
dose_2 <- teeth[teeth$dose == 2,]


# summaries and plots

summary(teeth)

qplot(dose,len,data=teeth,facets=.~supp)  # a quick scatterplot of each delivery method

plot1 <- ggplot(teeth, aes(supp,len))
plot1 <- plot1 + geom_boxplot()

plot2 <- ggplot(teeth, aes(dose, len))
plot2 <- plot2 + geom_boxplot()

mean(teeth$len)
var(teeth$len)
sd(teeth$len)

# t tests

supp_test_1 <- t.test(teeth$len, supp_vc$len, paired=FALSE)$conf.int
supp_test_1

supp_test_2 <- t.test(teeth$len, supp_oj$len, paired=FALSE)$conf.int
supp_test_2

dose_test_1<- t.test(teeth$len, dose_05$len, paired=FALSE)$conf.int
dose_test_1

dose_test_2 <- t.test(teeth$len, dose_1$len, paired=FALSE)$conf.int
dose_test_2

dose_test_3 <- t.test(teeth$len, dose_2$len, paired=FALSE)$conf.int
dose_test_3
```