toothgrowth

Investigation of Tooth Growth vs Supplement/Dose Combinations

L. Chin 01/30/16

Overview

The ToothGrowth dataset in R was studied to determine the effects of using orange juice (OJ) or vitamin C (VC) as a supplement (in various doses) to aid tooth growth.

```
data(ToothGrowth)
tg<-ToothGrowth
tg$dose <- factor(tg$dose)
library(ggplot2)
```

Exploratory Analysis

The first step in the analysis is to visually explore the relationship of various factors to growth in tooth length. The results of the exploratory phase are used to determine which scenarios require further investigation using t-tests. The visual exploration uses boxplots to show the distribution of tooth length for various combinations of supp and dose.

For those not familiar with boxplots, median values for the distribution are marked by the line in the center of the box, the upper and lower boundaries of box show the start of the top quartile and bottom quartile. The end of the lines at the top and bottom of the box show the maxium and minimum values in the distribution, with outliars appearing as dots above or below the end of the lines.

Tooth Length vs. Supp (OJ/VC)

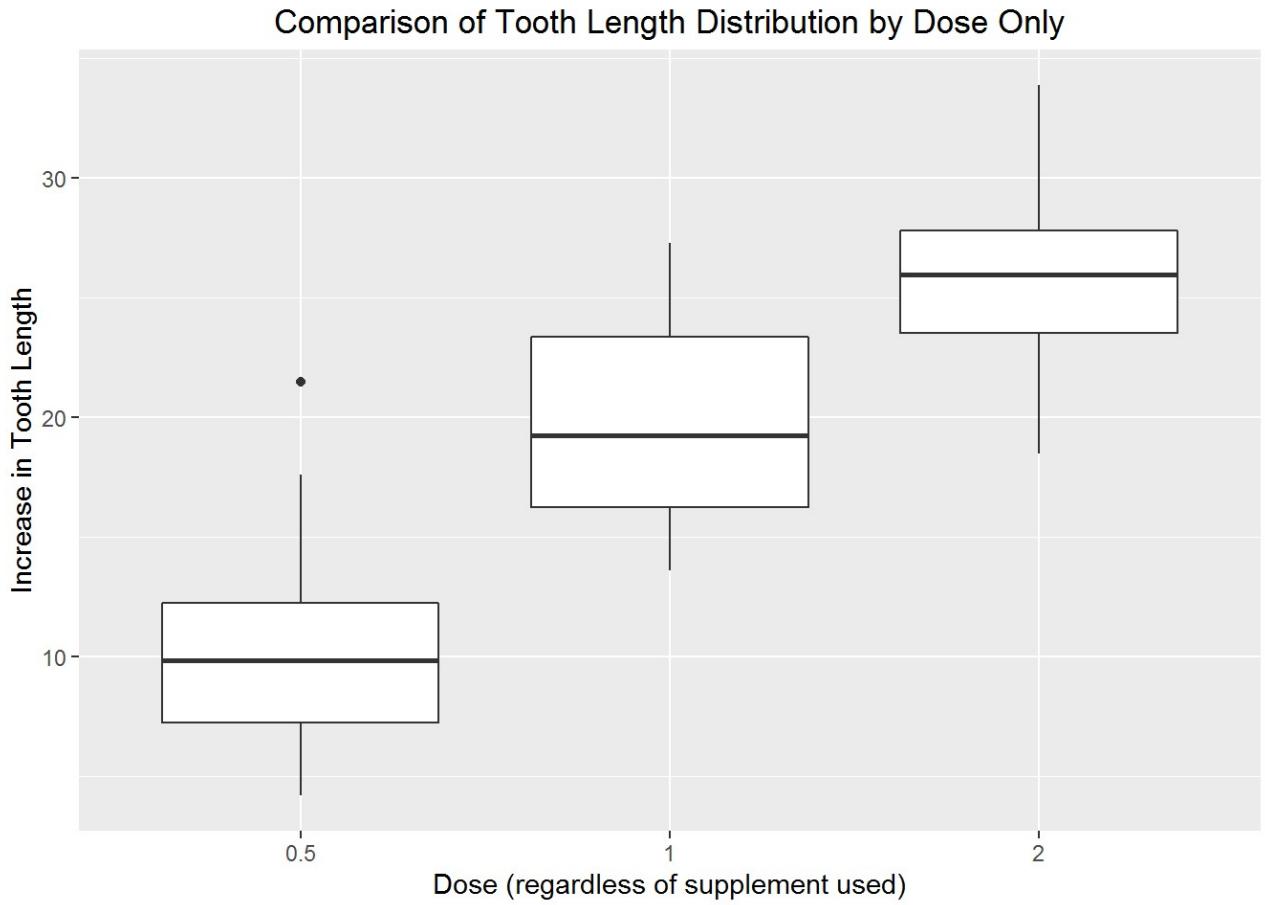
The chart below shows the distribution of tooth length vs the supplement used (OJ or VC). This is regardless of the dose used. The boxplot shows that without using any other distinguishing factors, there is an overlap between the range of tooth growth experienced by subjects using OJ vs VC. However, the median values differ, with OJ skewed toward a higher value overall, so it may still be possible to distinguish between the two groups.

```
# compares the distribution of tooth length by supp, regardless of dose
g<-ggplot(tg, aes(x=supp, y=len))
g<-g+geom_boxplot(aes(fill=supp)) + ggtitle("Comparison of Tooth Length Distrib
ution by Supp Only") + xlab("Supplement Used") + ylab("Increase in Tooth Lengt
h")
print(g)
```

Tooth Length vs Dose (Regardless of Supp)

The chart below shows the increase in tooth length vs the dose of a supplement given, regardless of the supp chosen. This clearly shows that by increasing the supplement dose, one can increase tooth growth, regardless of whether one uses OJ or VC.

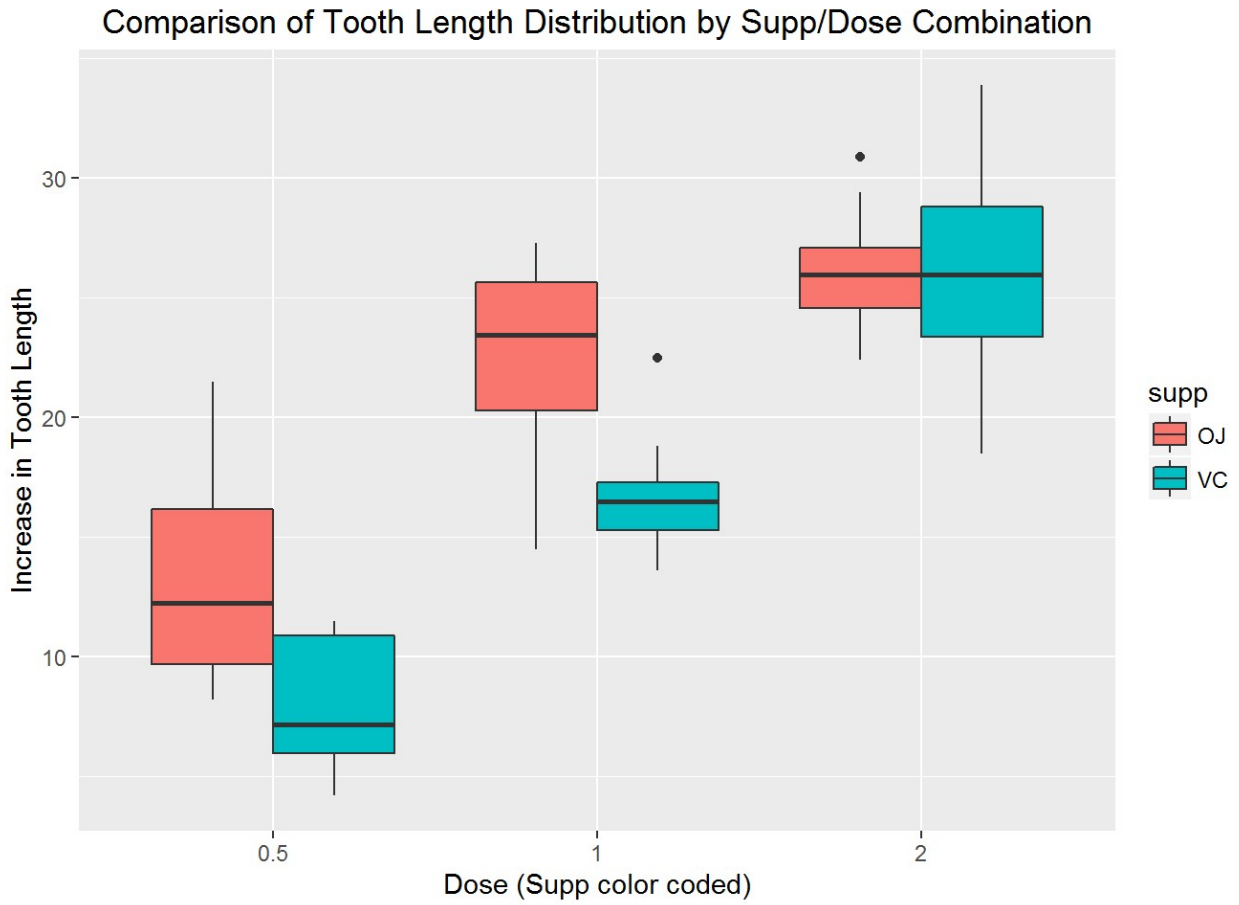
```
g<-ggplot(tg, aes(x=dose, y=len))
g<-g+geom_boxplot()+ggtitle("Comparison of Tooth Length Distribution by Dose On
ly") + xlab("Dose (regardless of supplement used)") + ylab("Increase in Tooth L
ength")
print(g)
```



Tooth Length vs Supp/Dose combination

The following chart explores at each dose level, whether OJ or VC produces more tooth growth. Here one can see that OJ clearly produces a signficant advantage, except at the highest dose, where both OJ and VC produce similar growth rates.

```
# compares the distribution of tooth length by dose, color coded by supp.  Allo
ws comparison
# of OJ/VC for 0.5, 1, and 2
g<-ggplot(tg, aes(x=dose, y=len))
g<-g+geom_boxplot(aes(fill=supp))+ggtitle("Comparison of Tooth Length Distribut
ion by Supp/Dose Combination")+xlab("Dose (Supp color coded)")+ylab("Increase i
n Tooth Length")
print(g)
```



Summarizing the visual findings:

1. The two groups (based on supp only), have somewhat similar ranges of increased tooth growth. This scenario is worth exploring in a t test to determine if there is a statistical difference between using OJ vs VC.
2. Clearly using a grouping based on dose only (regardless of supp), shows that tooth growth increases. Since this is quite clear, no t test is required to make the claim that increasing dosage increases tooth growth
3. Determining whether OJ or VC is preferred at various dose levels requires further investigation, and a t test is explored to determine which supp is preferred at each dose.

T test and Confidence Interval for Different Scenarios

The t test scenarios below reference the scenarios summarized in the Exploratory Section. For the test, the null hypothesis is that the difference between the means of Group A and Group B is zero. The alternate hypothesis is that the difference in means is not zero. A two sample test is used t.test(x, y, paired=FALSE)

Scenario 1 - Group A and B Based on Supplement Only

```
# setup the individual groups for further study
vc05<-ToothGrowth[1:10,]
oj05<-ToothGrowth[31:40,]
vc10<-ToothGrowth[11:20,]
oj10<-ToothGrowth[41:50,]
vc20<-ToothGrowth[21:30,]
oj20<-ToothGrowth[51:60,]

vc<-ToothGrowth[1:30,]
oj<-ToothGrowth[31:60,]
```

```
t.test(vc$len, oj$len, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc$len and oj$len
## t = -1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

The difference in means is -3.7. The test has a p-value of .06063, which means under the null hypothesis there is a 6% chance of getting a difference of -3.7, or worse. Since we normally use a two-sided 5% confidence interval, and 6% is > 5%, we would not reject the null hypothesis. Although we are in a marginal area since we are so close to the 5% mark. One can see that the 95% confidence interval for the alternative hypothesis includes the 0 mark. So under the alternative hypothesis, a difference of zero falls within the confidence interval.

Scenario 2 - Group A and B Based on Supplement and Dose=.5

```
t.test(vc05$len, oj05$len, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc05$len and oj05$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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

The p-value is very small at .006359. The difference obtained between these two groups is an unlikely event if the null hypothesis were true (<.6%). The 95% confidence interval for the alternative hypothesis does not include zero, so we reject the null hypothesis. There is a difference in means between Group A and B with Dose=.5. This seems consistent with the boxplot which shows a very small degree of overlap in the range of tooth growth experienced by these 2 groups

Scenario 3 - Group A and B based on Supplement and Dose=1

```
t.test(vc10$len, oj10$len, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc10$len and oj10$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:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

Tne p-value is even smaller for this group at .001038. The 95% confidence interval for the alternative hypothesis does not include zero (and is further from zero than with Dose=.5). This would mean we should reject the null hypothesis. There is a difference in means between Group A and Group B, with Dose=1. This is also consistent with the boxplot which shows very little overlap in the range of tooth growth experienced by these 2 groups

Scenario 4 - Group A and B based on Supplement and Dose=2

```
t.test(vc20$len, oj20$len, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc20$len and oj20$len
## t = 0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

The 95% confidence interval for the alternative hypothesis is almost centered at zero (i.e. it overlays on top of the distribution for the null hypothesis of difference in means=0). The p-value is .96, which means that the 96% of the values of the null hypothesis are worse than the difference obtained in this particular sample. In this case, you would not reject the null hypothesis.

Summarizing the T-Test Results

The t-test was run against several scenarios to determine the following:

1. Using OJ vs VC (regardless of dose) - one should accept the null hypothesis. The two groups A and B are sufficiently close to each other than the alternative hypothesis is not accepted. However, the p-value is right on the border (very close to alpha=.05), although p-value > alpha so we keep the null hypothesis.
2. Using a combination of supplement and dose. For Dose=.5 and Dose=1, OJ is preferable to VC as it produces more tooth growth (the null hypothesis was rejected for these 2 scenarios). For Dose=2, OJ and VC have indistinguishable results in tooth growth.