ToothGrowth Analysis

Ashly Yashchin

8/5/2017

Introduction

The following analysis is a basic exploration of the ToothGrowth dataset in R, exploring the effect of Vitamin C on tooth growth in Guinea Pigs.

There were 60 guinea pigs used in this sample. Each animal received 1 of 3 dose levels of Vitamin C (0.5,1,2) by one of two delivery methods (orange juice or ascorbic acid coded as VC).

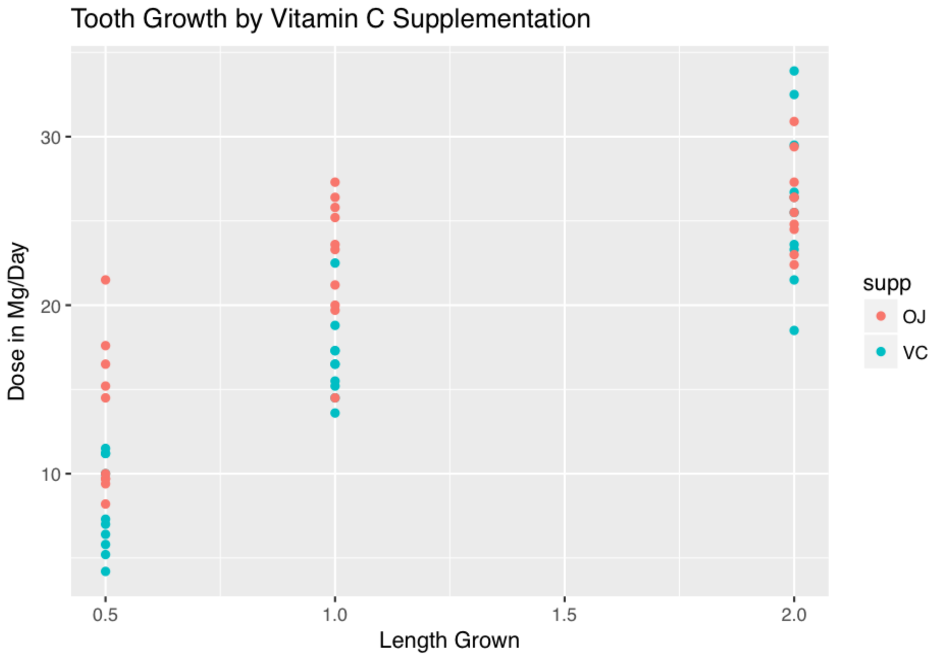
There are 60 observations and 3 variable columns: * len = the numeric tooth length * supp = the delivery method * dose = milligrams/day

Exploration

```
#First let's set up our workspace and load our data.
library(ggplot2)
library(dplyr)
data("ToothGrowth")
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
#Now, let's look at some exploratory plots.
ggplot(ToothGrowth,aes(x=dose,y=len))+geom_point(aes(color=supp))+ggtitle("Tooth Growth by Vitamin C Su
```

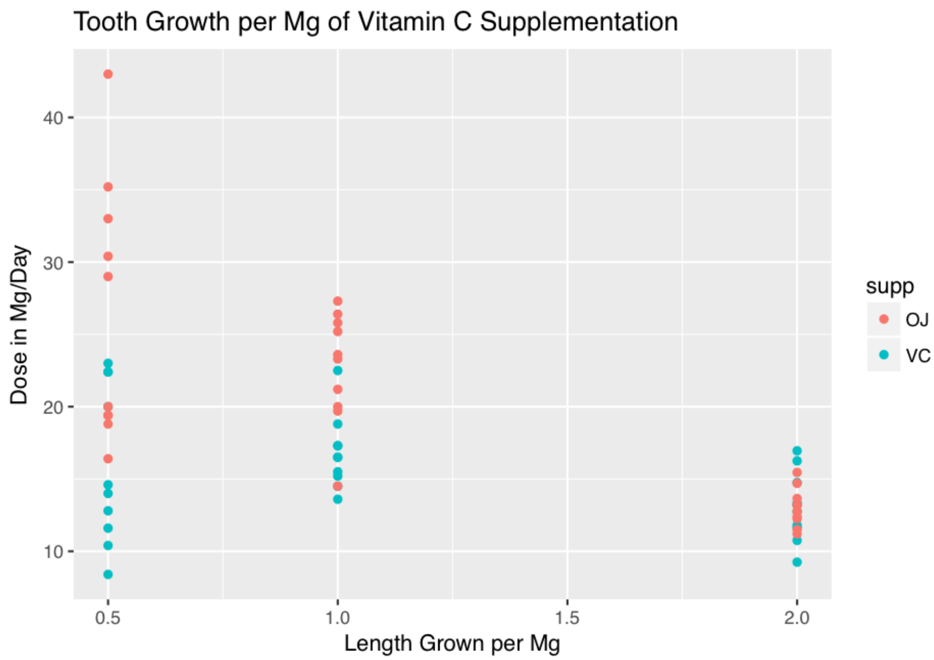


At first glance, it appears, that the Vitamin C administered in the form of OJ, resulted in greater growth, at least at the 0.5mg and 1.0mg levels. However, when we get to 2.0mg, it seems less clear.

Let's add another column to calculate the length grown per mg administered.

```
ToothGrowth <- mutate(ToothGrowth,len_mg=len/dose)

#Let's plot again.
ggplot(ToothGrowth,aes(x=dose,y=len_mg))+geom_point(aes(color=supp))+ggtitle("Tooth Growth per Mg of Vi
```



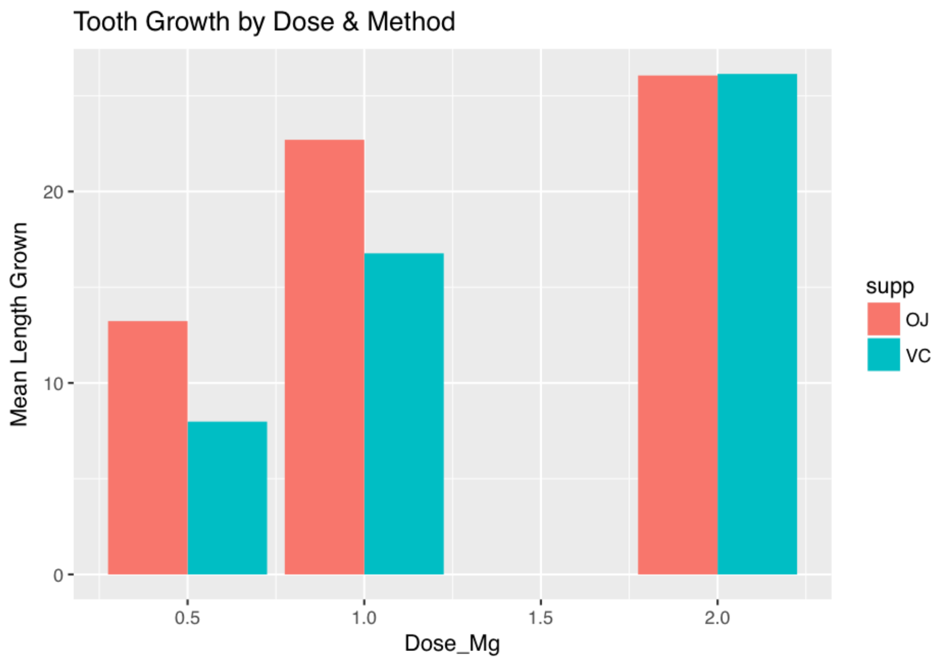
What's interesting here is we see an inverse of our first plot. In the first plot it's clear that the higher the dose given, the greater the tooth growth. However, when we look at it on a growth/mg basis, we see that the effect lessens at higher doses. Each subsequent 0.5 mg is not as effective as the first.

We also see that the variability is very high at the 0.5 dose and gets smaller as we get closer to 2.0.

Let's look at the supplementation types side by side.

```
#We need to group our data by dosage and then by supplementation type.
ToothGrowth <- group_by(ToothGrowth,dose,supp)

#Now, let's compare the mean length grown.
TGMeans <- summarize(ToothGrowth,mean(len),var(len),sd(len)) %>% rename(mean_len=`mean(len)`,var_len=`va
ggplot(TGMeans,aes(x=dose,y=mean_len,fill=supp))+geom_bar(position="dodge",stat="identity")+ggtitle("To
```




```r
print(TGMeans)
```

```
## Source: local data frame [6 x 5]
## Groups: dose [3]
##
## # A tibble: 6 x 5
##    dose   supp mean_len   var_len   sd_len
##   <dbl> <fctr>    <dbl>     <dbl>    <dbl>
## 1   0.5     OJ    13.23 19.889000 4.459709
## 2   0.5     VC     7.98  7.544000 2.746634
## 3   1.0     OJ    22.70 15.295556 3.910953
## 4   1.0     VC    16.77  6.326778 2.515309
## 5   2.0     OJ    26.06  7.049333 2.655058
## 6   2.0     VC    26.14 23.018222 4.797731
```

Testing

Hypothesis

My hypothesis is that, at doses of 1.0mg or less, OJ, is more effective than VC. At doses of 2.0 a clear winner cannot be determined.

In order to determine if this hypothesis is true we need to test at each dosage level.

```r
#Let's break out our data by dosage level.
OJDose.5 <- ToothGrowth$len[31:40]
```

```
OJDose1.0 <- ToothGrowth$len[41:50]
OJDose2.0 <- ToothGrowth$len[51:60]
VCDose.5 <- ToothGrowth$len[1:10]
VCDose1.0 <- ToothGrowth$len[11:20]
VCDose2.0 <- ToothGrowth$len[21:30]
```

Assumptions

We will assume variance is unequal (that seems pretty clear from our exploratory graphs). We will use a 95% confidence interval for our test.

The data is not paired as separate guinea pigs were used for each observation. None received multiple doses.

T-Tests

```
#0.5 Dose
#Null Hypothesis = mean(OJDose.5)==mean(VCDose.5)
#Alternative Hypothese = mean(OJDose.5) > mean(VCDose.5)
t.test(OJDose.5,VCDose.5,paired=FALSE,var.equal = FALSE,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJDose.5 and VCDose.5
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
#REJECT the Null Hypothesis

#1.0 Dose
#Null Hypothesis = mean(OJDose1.0)==mean(VCDose1.0)
#Alternative Hypothese = mean(OJDose1.0) > mean(VCDose1.0)
t.test(OJDose1.0,VCDose1.0,paired=FALSE,var.equal = FALSE,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJDose1.0 and VCDose1.0
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
#REJECT the Null Hypothesis
```

```r
#2.0 Dose
#Null Hypothesis = mean(OJDose2.0)==mean(VCDose2.0)
#Alternative Hypothese = mean(OJDose2.0) > mean(VCDose2.0)
t.test(OJDose2.0,VCDose2.0,paired=FALSE,var.equal = FALSE,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJDose2.0 and VCDose2.0
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

```r
#FAIL TO REJECT the Null Hypothesis
```

Results

Based on the above t.tests, it appaers that our hypothesis was correct. With p-values way less than 5% on the 0.5 mg and 1.0 mg dose level tests, we can say, with confidence, that OJ is the more effective measure of supplementation. However, when we get to the 2.0 dose level, we can not confidently say the same, as a p-value of 0.5181 means that 51.8% of the time, the results of OJ supplementation might equate to less growth than VC supplementation.

Even with these results, with a sample size of only 10 per variable, I would recommend a larger study to confirm these findings.