ToothGrowth Analytical Report

Dwight Shackelford

July 28, 2016

Tooth Growth Analytical Report

Overview: Analysis of The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Description

The underlying data tracks the response of the length of odontoblasts (cells responsible for tooth growth) to ascorbic acid supplementation, in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice (OJ) or ascorbic acid, a pure form of vitamin C (VC).

```
library(datasets)
data("ToothGrowth")

str(ToothGrowth)

## '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 ...
```

From the help(ToothGrowth) R function, the data is described as follows: A data frame with 60 observations on 3 variables.

[1] len numeric Tooth length

[2] supp factor Supplement type (VC or OJ).

[3] dose numeric Dose in milligrams/day

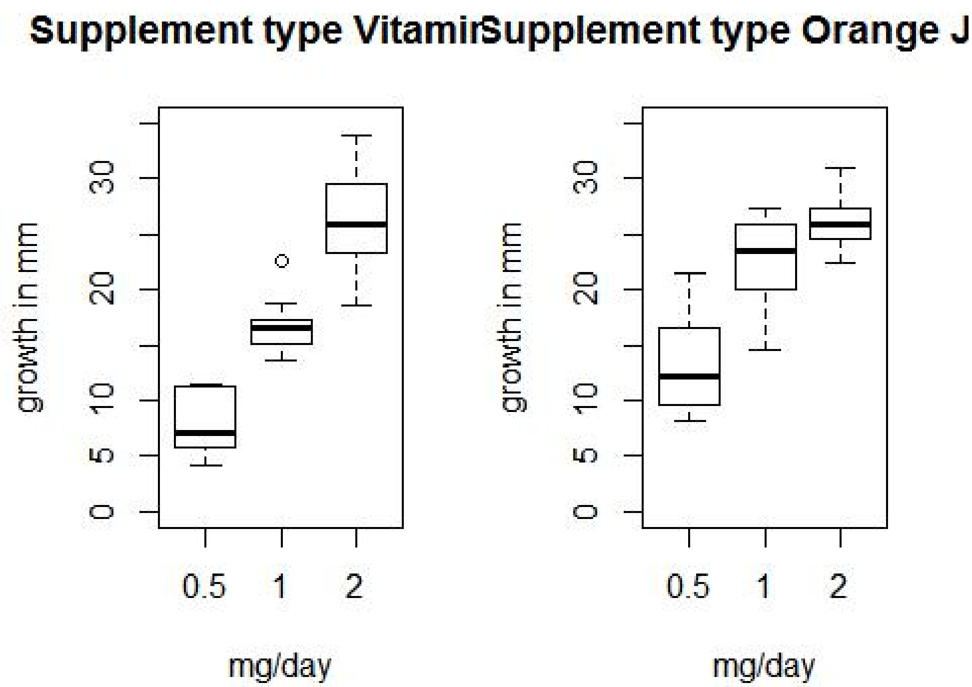
Analytical Plots

The following pair of plots shows the comparison of length of tooth growth as related to daily dosage.

The two methods of supplementation, Ascorbic Acid (Vitamin C or VC in the table), and Orange Juice (OJ in the table) are shown side by side in the two box plots below.

```
par(mfcol=c(1,2))
boxplot(len~dose, data=subset(ToothGrowth,supp=="VC"),xlab="mg/day",yla
b="growth in mm", ylim=c(0,35))
```

```
title("Supplement type Vitamin C")
boxplot(len~dose, data=subset(ToothGrowth,supp=="OJ"),xlab="mg/day",yla
b="growth in mm", ylim=c(0,35))
title("Supplement type Orange Juice")
```



Just visually looking at the plots, There is generally more positive results using the orange juice supplement over the vitamin C.

Since the values for 2.0mg/day are so close on the plot scales, I've broken them out here:

Variance Calcs

```
## var for OJ at 2.0
var(subset(ToothGrowth, supp=="OJ" & dose==2.0, select=c(len))$len)

## [1] 7.049333

## var for VC at 2.0
var(subset(ToothGrowth, supp=="VC" & dose==2.0, select=c(len))$len)

## [1] 23.01822
```

Variances are significantly different, so have to use different type of t test, i.e. var.equal=FALSE

Median and Mean calcs for 2.0mg/day:

```
## median for VC at 2.0mg/day
median(subset(ToothGrowth, supp=="VC" & dose==2.0, select=c(len))$len)

## [1] 25.95

## median for OJ at 2.0mg/day
median(subset(ToothGrowth, supp=="OJ" & dose==2.0, select=c(len))$len)

## [1] 25.95

## mean for VC at 2.0mg/day
mean(subset(ToothGrowth, supp=="VC" & dose==2.0, select=c(len))$len)

## [1] 26.14

## mean for OJ at 2.0mg/day
mean(subset(ToothGrowth, supp=="OJ" & dose==2.0, select=c(len))$len)

## [1] 26.06
```

Median and mean are very close on the 2.0mg/day.

Now to try some confidence intervals.

Confidence Interval Testing of Results

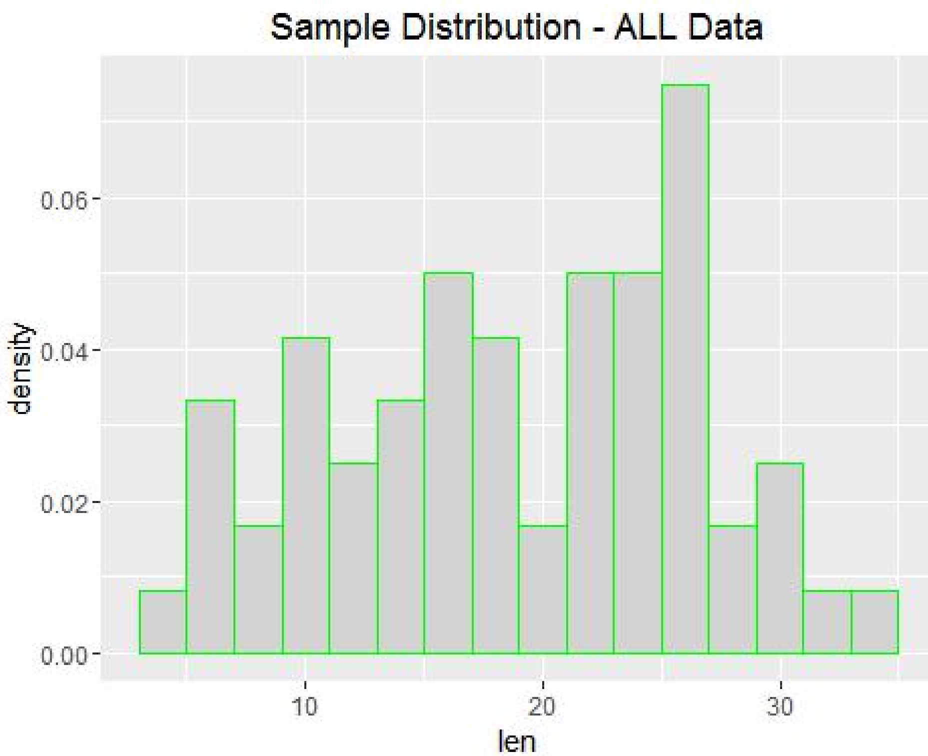
Lets check to see of the data is mound shaped and symmetrical. If so, we can use t.tests to analyze.

First set up the table for 2.0mg/day dosage levels for t.test

```
## set up 2.0mg/day table
TG_20<-subset(ToothGrowth, dose==2.0 & supp=="OJ", select=c(len))
TG_20<-cbind(TG_20,subset(ToothGrowth, dose==2.0 & supp=="VC", select=c
(len)))
colnames(TG_20)<-c("OJ","VC")
```

Now we check for symmetrical and normal distribution

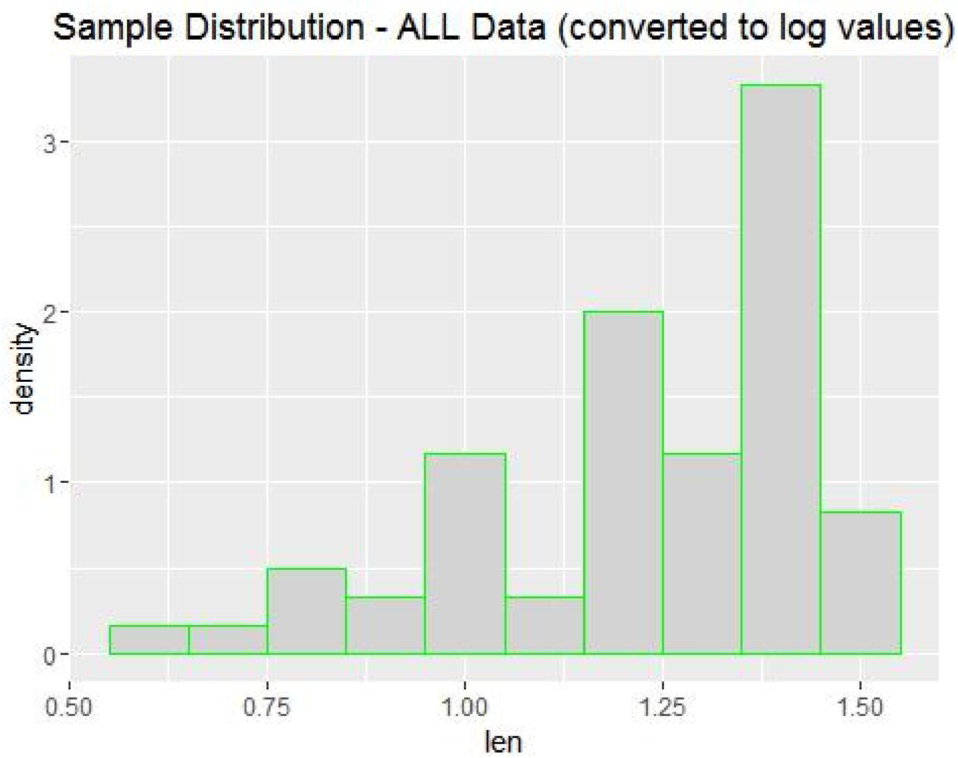
```
library(ggplot2)
g<-ggplot(data=data.frame(ToothGrowth), aes(x=len))+ geom_histogram(bin
width=2, aes(y=..density..), fill="lightgrey", color="green")+ggtitle("
Sample Distribution - ALL Data")
print(g)
```



Data does not look symmetrical or normal.

Now we convert len to log value, to see if normal distribution traits arise.

```
ToothGrowth$len<-log10(ToothGrowth$len)
library(ggplot2)
g<-ggplot(data=data.frame(ToothGrowth), aes(x=len))+ geom_histogram(bin
width=.1, aes(y=..density..), fill="lightgrey", color="green")+ggtitle
("Sample Distribution - ALL Data (converted to log values)")
print(g)
```



Summary of all data is non-symmetrical and non-normal.

Graphs of each permutation of Supplement and Dosage (not shown here due to space limitations) showed similar non-symmetrical traits.

Sample size of 10 (per supplement) for 2.0mg/day is under 30, so t.test might not be valid to use. However, it was dictated by the instructor.

Confidence Interval tests

```
t.test(TG_20$OJ,TG_20$VC, paired=FALSE, var.equal=FALSE)

##
##  Welch Two Sample t-test
##
## data:  TG_20$OJ and TG_20$VC
## 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:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Conclusions

I have reservations that t.tests and confidence intervals can correctly be used on this data, due to small sample size, non-symmetrical data, and non-normal distribution. However, we were instructed to use this, so I have fulfilled that requirement.

Given the caveats, the analysis results are as follows:

For dosages below 2.0mg per day, the type of supplement matters regarding tooth growth, with Orange Juice being more effective that Vitamin C.

For dosages of 2.0mg per day, the t.test has 0.0 within the confidence interval, which says one can't rule out that the mean difference is 0. Therefore, the type of supplement used at 2.0mg per day does not have a statistical difference between the outcomes.