ToothGrowth Analysis

Alex Lee

10/16/2019

Overview

The goal of this project is to analyze the average tooth growth given a two varities of treatments and varying dosages. We will start by loading the ToothGrowth data set and perform initial data analysis to get a good understanding of the provided data set. Next, we will perform additional tests to determine if the samples provided give enough guidance as to the effectiveness of the treatment.

Loading Required Libraries and Data Set

This analysis will require the ggplot2 library and the command to load the ToothGrowth data set:

```
library(ggplot2)
data("ToothGrowth")
```

Once the data is loaded into R, we will need to take a close examination of the data provided by the data set. We will first take a look at the column names:

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

We can see that this is a very simple 3 column table consistig of Length, Supplement and Dosage level. Next we'll take a look at how many different supplements were part of this treatment:

```
levels(ToothGrowth$supp)
```

```
## [1] "OJ" "VC"
```

And it appears that we have two different types of supplements used in the test - OJ and VC. Now we want to take a look at the various treatment dosages and get a rough estimate on how big this sample sets are:

```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```
table(ToothGrowth$supp)

##
## OJ VC
## 30 30
```

The data seems to indicate that there are three different dosage levels - 0.5, 1 and 2. Also, there are a total of 60 samples provided in this data set, split up into 30 samples for each dosage type.

Now to round out initial data analysis, we will extract the summary statistical information of the data set:

```
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
```

Based on the quartiles for length, there appears to be some growth occuring among the test as a whole. Additional data ananlysis will be needed to support the findings of these tests. Additionally, we can see there are correlations between dosage and length as well:

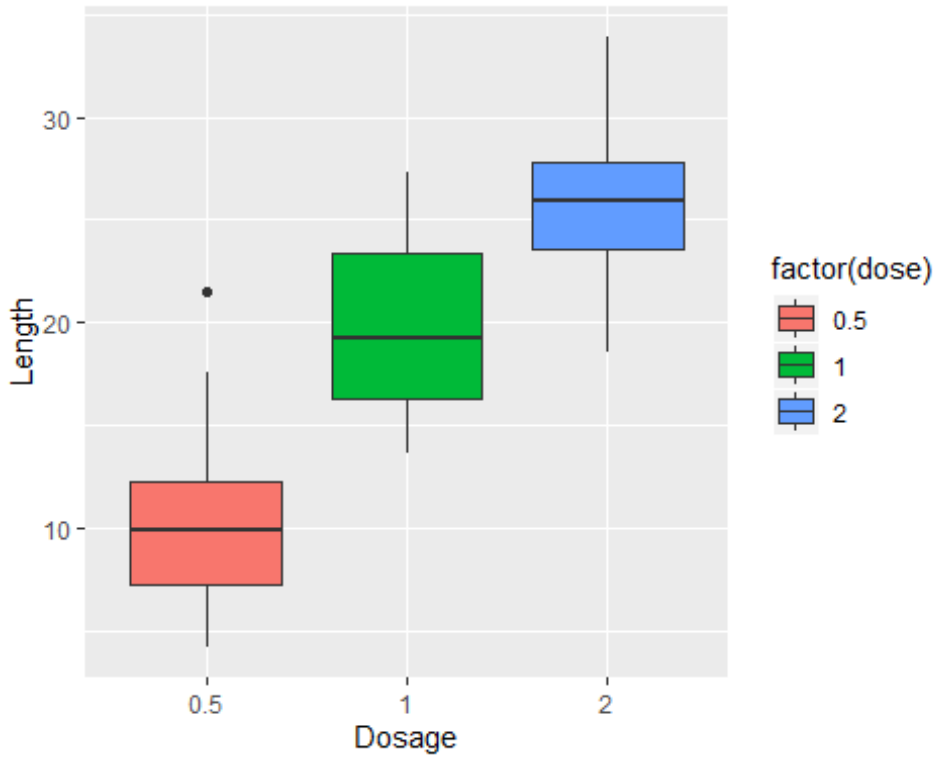
```
cor(ToothGrowth[sapply(ToothGrowth, is.numeric)])

##            len      dose
## len  1.0000000 0.8026913
## dose 0.8026913 1.0000000
```

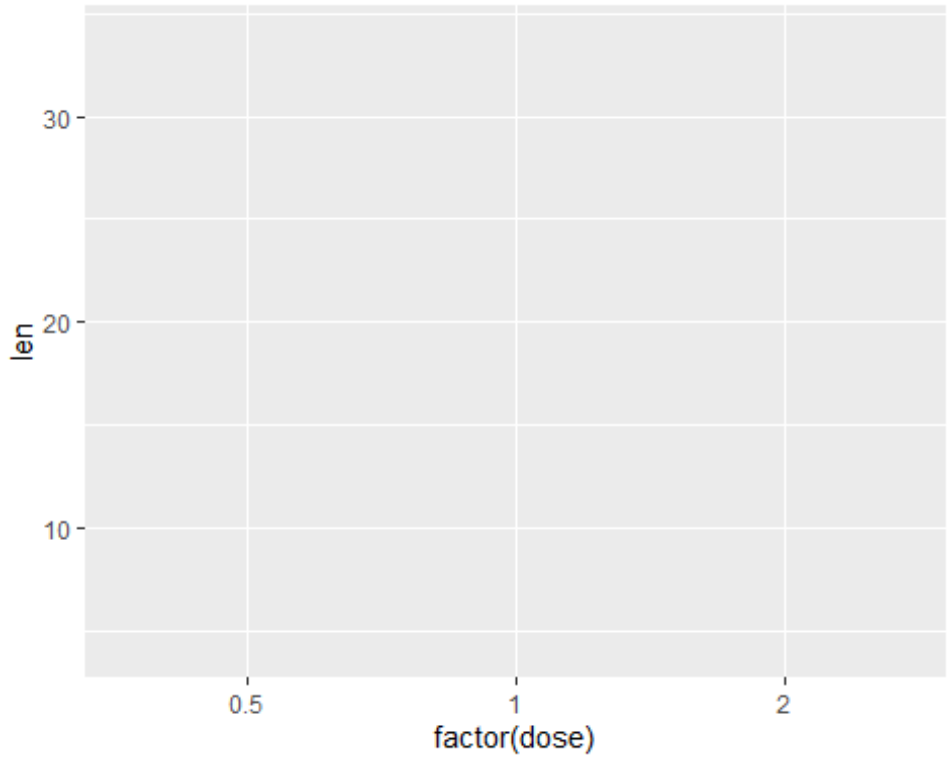
Visualizing the Data

Next, we will need to visualize the data provided by the ToothGrowth data set. The best plot to use in this instance is the boxplot with colorizations for the varying dosage types.

```
p1 <- ggplot(ToothGrowth, aes(x =factor(dose), y = len, fill = factor(dose)))
p1 + geom_boxplot() + xlab("Dosage") + ylab("Length")
```


30
20
10
Length
0.5
1
2
Dosage
factor(dose)
0.5
1
2

```
p1
```


30
20
10
len
0.5
1
2
factor(dose)

Based on the boxplots, its clear that the larger the dose, the larger the tooth growth. It also appears that the rate of tooth growth seems to be between 0.5 and 1 where the median almost doubles - the same cannot be said when dosage levels g from 1 to 2. Also, dosage at 1 seems to have the largest ranges of growth.

Confidence Intervals, Significance and T Tests

We will use R t.test function to examine the difference in means for the two dosage samples provided by the ToothGrowth data set. The null hypothesis assumes that regardless of the supplement type, the tooth growth lengths will be the same or the mean difference is 0. The alternative hypothesis will be the supplements does have an effect of tooth growth. We will also assume two-sided tests, non-equal variances and the two test groups are not paired:

```
t.test(len~supp, mu=0, alt="two.sided", var.eq=F, paired=F, , conf=0.95,
data=ToothGrowth)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Based on the p-value (0.06) being less than confidence interval 0.95 (alpha) I can safely reject the null hypothesis and assume the alternative hypothesis is true. This is also supprted by the fact the true difference in means between the two dosage group are not 0.

We will now split our tests based on the dosage leve. We will first start by subsetting the ToothGrowth data set by dosage amount:

```
dose0_5 <- subset(ToothGrowth, dose == 0.5)
dose1   <- subset(ToothGrowth, dose == 1)
dose2   <- subset(ToothGrowth, dose == 2)
```

We will now run t.test 3 times since t.test only accepts two populations at a time.

```
t.test(len~supp, mu=0, alt="two.sided", var.eq=F, paired=F, , conf=0.95,
data=dose0_5)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp, mu=0, alt="two.sided", var.eq=F, paired=F, ,
conf=0.95,data=dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len~supp, mu=0, alt="two.sided", var.eq=F, paired=F, ,
conf=0.95,data=dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##            26.06            26.14
```

The null hypothesis assumes that there is 0 tooth growth regardless of the dosage size while the alternative hypothesis assumes greater than 0 growth. We assume the same parameters as before.

For all three tests, while the null hypothesis can be rejected for all three dosage types, the dosage2 tests seems to provide the greatest percentage (or p-value) amnong the group.

Next we will compare the effectiveness of OJ and VC and this will require subsetting the data into separate data frames:

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

We will now run t tests to determine if one supplement provides better results than the other at a high level:

```
t.test(oj$len, vc$len, var.eq=T)
```

```
##
##  Two Sample t-test
##
## data:  oj$len and vc$len
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

In this case, we do see that OJ seems to have the better overall result than VC. The p-value at 6% is less than our alpha and so we can reject the null hypothesis that states equal results for the two supplement types.

We will now compare dosage levels and supplement type and determine which of the combinations are more effective than the other. We first start by subsetting the oj and vc data set into dosage level tables:

```
oj0_5 <- subset(oj, dose==0.5)
oj1 <- subset(oj, dose==1)
oj2 <- subset(oj, dose==2)

vc0_5 <- subset(oj, dose==0.5)
vc1 <- subset(oj, dose==0.5)
vc2 <- subset(oj, dose==0.5)
```

We will assume for each t test that the null hypothesis is that both show the same effectiveness (eg, lengths) where the alternative would be unequal lengths. This will determine at what combination of dosage level and supplement type yields the most effective combination:

```
t.test(oj0_5$len, vc0_5$len, var.eq=T)

##
##  Two Sample t-test
##
## data:  oj0_5$len and vc0_5$len
## t = 0, df = 18, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.190168  4.190168
## sample estimates:
## mean of x mean of y
##     13.23     13.23

t.test(oj1$len, vc1$len, var.eq=T)

##
##  Two Sample t-test
```

```
##
## data:  oj1$len and vc1$len
## t = 5.0486, df = 18, p-value = 8.358e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.529186 13.410814
## sample estimates:
## mean of x mean of y
##     22.70     13.23
```

```
t.test(oj2$len, vc2$len, var.eq=T)
```

```
##
##  Two Sample t-test
##
## data:  oj2$len and vc2$len
## t = 7.817, df = 18, p-value = 3.402e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   9.381777 16.278223
## sample estimates:
## mean of x mean of y
##     26.06     13.23
```

Conclusions

Based on the final series of tests, it is concluded that the supplement OJ at lower dosage levels (0.5, 1) yields better reuslts than VC at the same leve. This is shown by a high T-value that is greater than the value represented in the statistics t-table.

In short, tooth length growth yields the greatest result based on dosage amount. Under lower dosage levels, OJ is more effective supplement than VC.