Statistical Inference Course Project Par 2

MLauden

February 22, 2017

Overview

The ToothGrowth dataset in R shows contains data on a study involving the influence of Vitamin C on tooth growth in guinea pigs. This analysis uses R to explore the contents of this data. Based on the initial analysis, a set of graphs are plotted to provide a qualitative understanding of possible outcomes in the study. Finally, hypothesis testing is used to provide more rigororus conclusions about the effects observed in the data.

Summary of ToothGrowth Data

The first step is to determine what the dataset is representing. The easiest way to do this is using the "?ToothGrowth" command. According to the help instructions, the data set represents, "The Effect of Vitamin C on Tooth Growth in Guinea Pigs".

Let's take a look at the first few rows of the data followed by the meaning of the row headings, and then a brief summary of the data:

```
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
```

```
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 ...
```

```
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  
```

It appears the units of tooth length are not available from this data. A link to the original paper is available, but the document is behind a paywall. This information would be useful to have for any practical application. However, the exact length is not required for the current exercise.

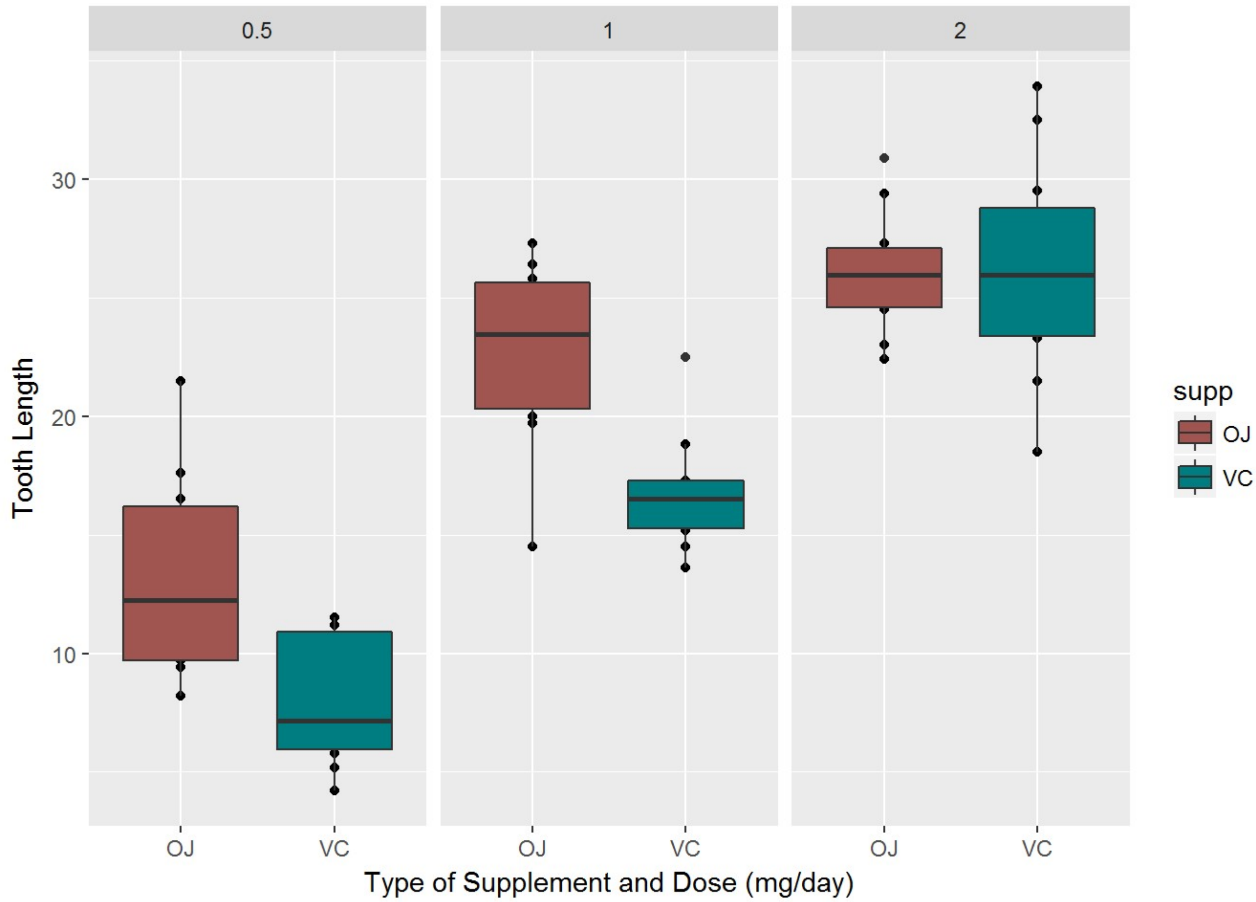
Now let's try to graph the data based on the various factors to see if anything stands out. Qualitatively, there appears to be an increase in tooth length with increasing dosage concentration of Vitamin C. There also appears to be a stronger influence on Tooth Length using OJ than ascorbic acid(VC) for the lower dosage amounts, but not the higher. Statistical analysis is needed to get a better idea of whether these effects are real.

```
require(ggplot2)
```

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

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
qplot(supp,len, data=ToothGrowth,
      #geom="boxplot",
      ylab="Tooth Length",
      xlab="Type of Supplement and Dose (mg/day)",
      facets=~dose
) + geom_boxplot(aes(fill = supp)) + scale_fill_hue(c=55, l=45)
```



Hypothesis Testing:

Supplement Type

Let the null hypothesis be that both supplement types influenced growth rate equally: H0: len(OJ)=len (VC) The alternate hypothesis will be that the OJ supplement had a larger affect on growth. Ha: len(OJ) >len(VC)

Subset the data by the supplement type:

```
#Subset the data
lenOJ<-subset(ToothGrowth, supp=="OJ",select = len)
lenVC<-subset(ToothGrowth, supp=="VC", select =len)

#Use T-Test One tailed with unequal variance and a 95% confidence interval (ass
umes a normal distribution)
t.test(lenOJ,lenVC,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOJ and lenVC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The T-Test shows a P-value of 0.03, which is below 0.05 indicating that the null hypothesis can be rejected. Therefore, there is an effect caused by the supplement type suggesting that OJ has a larger influence on tooth growth than VC. Note: That this test was based off of the initial data exploration that suggested OJ increased growth more than VC. If there was no assumption about which method had a larger effect, a two-sided T-Test would be used. In that case, the P-value would be double 0.06063, and one would not be able to reject the null hypothesis. To better understand the relationship between the supplement types, the next section will look at the supplement effect vs. the different dosage levels. The data show that there is a significant effect at 0.5 and 1 mg/day dosage, but not at the highest level (2mg/day). At the high dosage, the null hypothesis cannot be rejected p-value=0.5181, which is greater than the 0.05 level of significance selected for the test.

```
#Subset the data
lenOJ0.5<-subset(ToothGrowth, supp=="OJ" & dose==0.5, select = len)
lenVC0.5<-subset(ToothGrowth, supp=="VC" & dose==0.5, select =len)
lenOJ1<-subset(ToothGrowth, supp=="OJ" & dose==1, select = len)
lenVC1<-subset(ToothGrowth, supp=="VC" & dose==1, select =len)
lenOJ2<-subset(ToothGrowth, supp=="OJ" & dose==2, select = len)
lenVC2<-subset(ToothGrowth, supp=="VC" & dose==2, select =len)

#Use T-Test One tailed with unequal variance and a 95% confidence interval (ass
umes a normal distribution)

#Test the effect of supplement type at the 0.5 mg/day dosage
t.test(lenOJ0.5,lenVC0.5,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOJ0.5 and lenVC0.5
## 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
```

```
#Test the effect of supplement type at the 1 mg/day dosage
t.test(lenOJ1,lenVC1,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOJ1 and lenVC1
## 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
```

```
#Test the effect of supplement type at the 2 mg/day dosage
t.test(lenOJ2,lenVC2,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOJ2 and lenVC2
## 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
```

Dosage

Let the null hypothesis be that the dosage types influenced growth rate equally: H0: len(OJ)=len(VC) The alternate hypothesis be that the dosage influences the len over the range used in the study. Ha: len(OJ)>len(VC)

Subset the data by the supplement type:

```
#Subset the data
lenLowD<-subset(ToothGrowth, dose==0.5,select = len)
lenMidD<-subset(ToothGrowth, dose==1, select =len)
lenHighD<-subset(ToothGrowth, dose==2, select =len)

#Use T-Test One tailed with unequal variance and a 95% confidence interval (ass
umes a normal distribution)
t.test(lenHighD,lenLowD,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenHighD and lenLowD
## t = 11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.27926      Inf
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

The T-Test shows a P-value of 2.199e-14, which is below 0.05 indicating that the null hypothesis can be rejected. Therefore, there is an effect caused by increasing the dosage from 0.5 to 2 mg/day. There are three levels of dosage, and only one was considered above. The next section will examine the possible effect of the intermediate levels of 0.5 to 1 and 1 to 2. In both of these cases, the T-Tests have a P value low enough to reject the null hypothesis. That is to say, increasing the dosage over the intervals used in the study has an observable effect.

```
#Test the 0.5 to 1 mg/day increase
t.test(lenMidD,lenLowD,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenMidD and lenLowD
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
#Test the 1 to 2 mg/day increase
t.test(lenHighD,lenMidD,alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  lenHighD and lenMidD
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Conclusions

Overall, this study suggests that vitamin C administered to guinea pigs can increase the length of odontoblasts. The two variables tested in the ToothGrowth dataset have significant effects. There is a difference in the performance of the supplement type at the two lower dosages. This suggests, at low dosages, OJ enhances tooth growth to a greater extent than VC, while the two supplements have a similar effect at a dosage of 2mg/day. The second effect is that increasing the dosage increases the extent of tooth growth for all dosage intervals used in the study irrespective of the form of the supplement.