Analysis of ToothGrowth Data

This work is done for the course project for Coursera Statistical Inference. The project uses the ToothGrowth dataset in R.

Introduction

Load the data into R and examine the file. Load ggplot library for plots.

```
library(ggplot2)
library(plyr)
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
```

```
str(ToothGrowth)
```

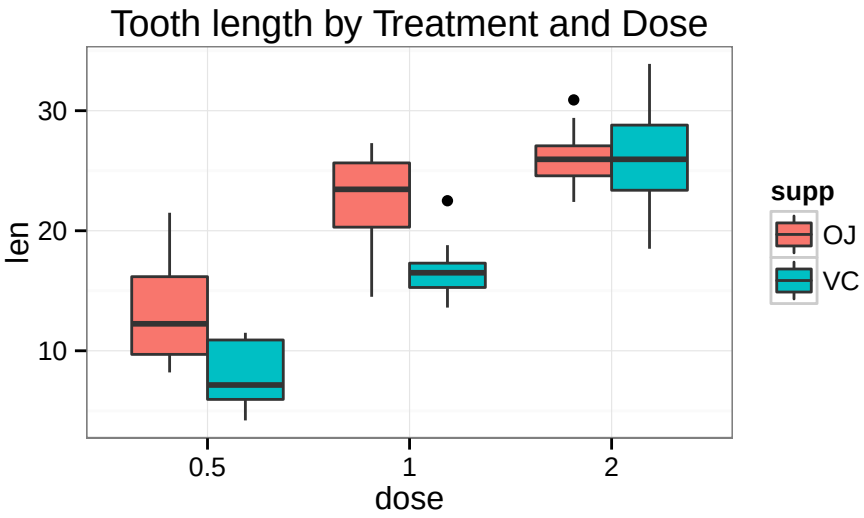
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Three variables that are len (numeric), supp (factor), and dose (numeric). But dose should be a factor, so convert it from a numeric to factor.

```
ToothGrowth$dose <-as.factor(ToothGrowth$dose)
```

Exploratory Data Analysis

Create a boxplot which will display the mean tooth length, the interquartile range, and total range for the data, by both dose group (0.5, 1, or 2), and separated by supplement group (OJ or VC). This will best show the potentially difference between the groups.



Dose of both types of supplement appear to have an effect on tooth length.

Data Summary

Summarize the tooth length data, finding the overall mean, as well as the mean by group (group being dose and supplement group).

```
### mean of columns
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```
### sum and mean of len by supp and dose
ddply(ToothGrowth, .(supp, dose), summarise, mean=mean(len), sum=sum(len))
```

```
##   supp dose  mean   sum
## 1   OJ  0.5 13.23 132.3
## 2   OJ    1 22.70 227.0
## 3   OJ    2 26.06 260.6
## 4   VC  0.5  7.98  79.8
## 5   VC    1 16.77 167.7
## 6   VC    2 26.14 261.4
```

Confidence intervals and hypothesis test

Test Tooth Length by Supplement

Run a t-test to assess if supplement leads to significant changes in tooth growth.

```
t.test(len ~ supp, paired = F, var.equal = F, 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
```

P-value is greater than 0.05, therefore we can not reject the null hypothesis that this is not difference between the two supplement groups on tooth length. We can further see evidence of this given the 95% confidence intervals contain 0.

Test Tooth Length by Dose

In order to examine the effects of dose on tooth length, we first need to create dummay variables for dose in order to perform a t-test (this is because dose has three levels instead of two).

```
dose12 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
dose13 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
dose23 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
```

Run t-tests comparing dose 0.5 to 1, 0.5 to 2 and 1 to 2.

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose12)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, paired = F, var.equal = F, data = dose13)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```r
t.test(len ~ dose, paired = F, var.equal = F, data = dose23)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Each of those t-test results i p-value below 0.05, which means we can reject the null hypothesis that there is no difference between the groups, and instead accept the alternative hypothesis. Furthermore, we can also see in the output that the 95 % confidence intervals for each t-test do not contain 0. Therefore it seems that dose is having an effect on tooth length.

Test Tooth Length by Dose and Supplement

Subset the data by dose.

```r
dose5 <- subset(ToothGrowth, dose == 0.5)
dose10 <- subset(ToothGrowth, dose == 1.0)
dose20 <- subset(ToothGrowth, dose == 2.0)
```

Run t-tests within each subset of dose for the relationship between tooth length and supplement.

```r
t.test(len ~ supp, paired = F, var.equal = F, data = dose5)
```

```
##
##  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
```

```r
t.test(len ~ supp, paired = F, var.equal = F, data = dose10)
```

```
##
##  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, paired = F, var.equal = F, data = dose20)
```

```
##
##  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
```

P-values were below 0.05 in the dose subests of 0.5 and 1.0, but not for 2.0. This means we can reject the null hypothesis for the for supplement type and dose effect tooth length together at doses 0.5 and 1.0 but not at 2.0.

Conclusions

We see that there are difference in tooth length by supplement at specific doses, but not all doses. We see for doses 0.5 and 1.0 OJ results in significantly ($p<0.05$) longer tooths than VC. However, this is not true for doses of 2.0 where there was not a statistically significant difference in tooth length between the groups.