Inferential Data Analysis on the ToothGrowth R Dataset

Juan Jose

Tuesday, March 17, 2015

```r
#Load any necessary datasets
library(datasets)

library(ggplot2)

#Store the ToothGrowth data frame into a variable, t, in case we do any
transformations
t = ToothGrowth
```

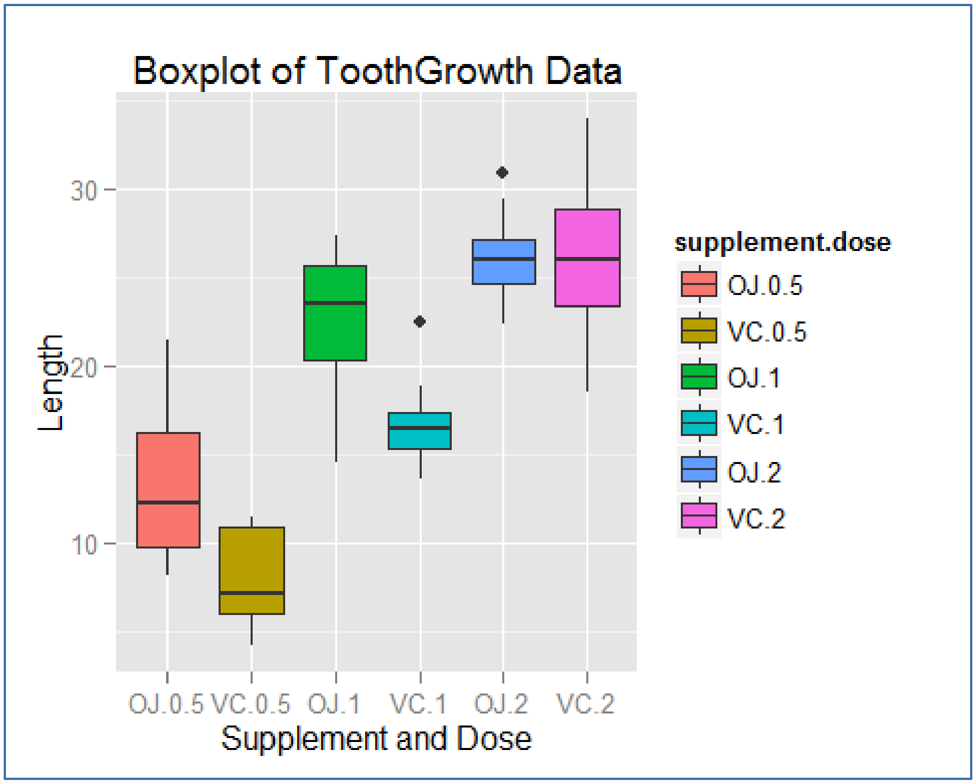
Exploratory Data Analysis

First we'll get a feel for what is in the ToothGrowth data frame.

```r
head(t)

##    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

summary(t)

##       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

unique(t$dose)

## [1] 0.5 1.0 2.0

unique(t$supp)

## [1] VC OJ
## Levels: OJ VC
```

We'll combine the dose and supp factors together and plot that against the length to get a better understanding of what is going on with the data.

```r
t$supplement.dose = interaction(t$supp, t$dose)

q = qplot(x=supplement.dose, y=len, data=t, geom="boxplot",
fill=supplement.dose, main="Boxplot of ToothGrowth Data",
ylab="Length", xlab="Supplement and Dose")

print(q)
```



Basic Summary Of the Data

From the resulting boxplot, we can tell that the the length of a tooth increases as the dosage of each supplement increases. It also appears to be the case that for dosages of 0.5 and 1, OJ causes teeth to be longer than VC, while for dosages of 2, there is no difference between the two supplements.

Comparison of Tooth Growth by Supplement

```r
supp.VC = t[which(t$supp=="VC"), 1]
supp.OJ = t[which(t$supp=="OJ"), 1]

var(supp.OJ)

## [1] 43.63344

var(supp.VC)

## [1] 68.32723
```

Because the variances are different, and the two groups are statistically indpendent we can call t.test with paired=F and var.equal=FALSE

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

```
##
##  Welch Two Sample t-test
##
## data:  supp.OJ and supp.VC
## 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 of x mean of y
##  20.66333  16.96333
```

Comparison of Tooth Growth by Dose

```
lowDose = t[which(t$dose==0.5), 1]
medDose = t[which(t$dose==1), 1]
highDose = t[which(t$dose==2), 1]
```

Test 1 vs 0.5

```
t.test(medDose, lowDose, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  medDose and lowDose
## 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:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Test 2 vs 1

```
t.test(highDose, medDose, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  highDose and medDose
## 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:
##  3.733519 8.996481
```

```
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Test 2 vs 0.5

```
t.test(highDose, lowDose, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  highDose and lowDose
## 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:
##  12.83383 18.15617
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

Conclusions and Assumptions

Based on the following assumptions:

- The t.tests on the dosage samples assume that the variances are different
- The are no confounding variables on the data
- The data are randomly selected and provide a representation of the entire population

Then we can conclude that:

- The supplement type does not affect the tooth length
- Dosage types do affect the tooth length, namely, the higher a dose, the larger the tooth length