

## Inferential Data Analysis on Tooth-Growth

Author: Byron.Bian

Date: Sunday, May 17, 2015

Overview: In this report, I will make a fundamental data exploration onto the dataset of Tooth-Growth and use confidence interval/hypothesis test to research the relationship between growth length and supp as well as dose

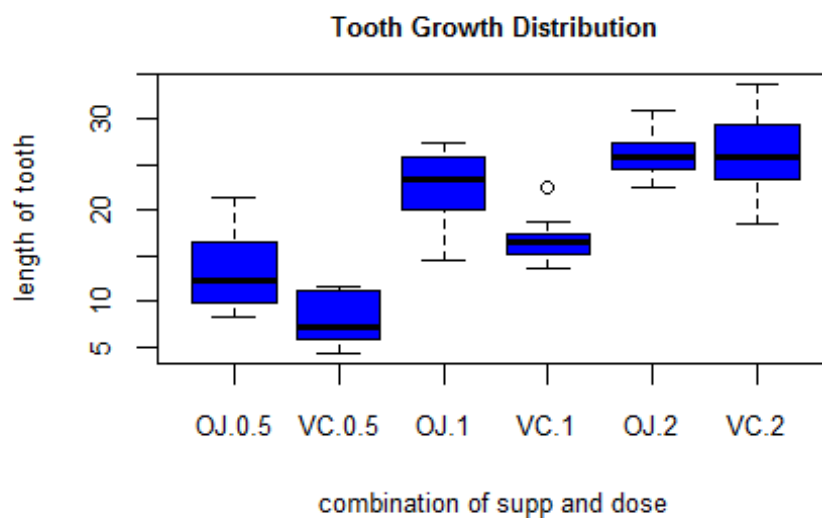
### Section I: Basic Exploratory Data Analysis

```
library(datasets); data(ToothGrowth); x<-ToothGrowth; summary(x)
```

```
##      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
```

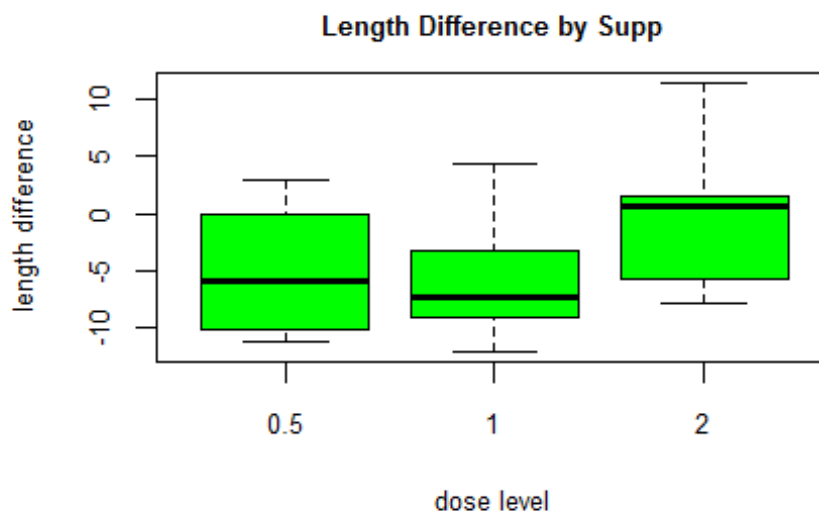
Next I can explore the data distribution via combinations by supp and dose

```
par(oma=c(0,0,0,0),mai=c(1,1,.5,.5)); x$dose<-as.factor(x$dose)
boxplot(len~supp+dose,data=x,notch=FALSE,varwidth=TRUE,col="blue",mai="Tooth Growth Distribution",
        xlab="combination of supp and dose", ylab="length of tooth",
        cex.axis=.8,cex.main=.8,cex.lab=.8)
```



From this plot, I found that distributions of different combinations are various, even for the same supp, different dose can lead to various mean level. Next I can explore the length difference between VC and OJ under the same dose level, in order to view the mean value level.

```
par(oma=c(0,0,0,0),mai=c(1,1,.5,.5));
y.vc<-subset(x,as.character(supp)=="VC"); y.oj<-subset(x,as.character(supp)=="OJ")
y<-cbind(y.vc[,c(3,1)], y.oj[,1]); names(y)[c(2,3)]<-c("len.vc","len.oj")
y$len.diff<-y$len.vc-y$len.oj;
boxplot(len.diff~dose,data=y,notch=FALSE,varwidth=TRUE,col="green",main="Length Difference by Supp",
        xlab="dose level", ylab="length difference",cex.axis=.8,cex.main=.8,cex.lab=.8)
```



```
rm(y.vc,y.oj)
```

We can see that mean values of difference are obviously not zero (except dose=2), so supp should be a significant dominant variable to tooth length

## Section II: Hypothesis test on growth length

Firstly, we can test the significance between different dose level under the same supp group. For example, we can test dose levels of .5 and 1 within VC. Because tooth growth data split by different supp or dose have the same size and are not independent, so we can use pair comparison, which is the assumption of test.

+H0: Difference in means is equal to 0

+Ha: Difference in means is not equal to 0

```
g1<-subset(x,as.character(supp)=="VC" & as.numeric(as.character(dose))==.5)$len
g2<-subset(x,as.character(supp)=="VC" & as.numeric(as.character(dose))==1)$len
```

```
e))==1)$len
t.test(g1,g2,paired=TRUE)$conf;t.test(g1,g2,paired=TRUE)$p.value

## [1] -12.030399 -5.549601
## attr(,"conf.level")
## [1] 0.95

## [1] 0.0001715165
```

We can see the p-value is far less than 5% and confidence interval doesn't include 0. Then we can reject  $H_0$  and accept  $H_a$ , means of different dose level are significantly not equal. Using the similar way, I found the differences are significant except the comparison that dose levels are 1 vs 2 within OJ (p.value=0.08384, failed to reject  $H_0$ ). Secondly, I can test difference in means between VC and OJ within the same dose level.

+ $H_0$ : Difference in means is equal to 0 + $H_a$ : Difference in means is not equal to 0

```
y.cmp<-subset(y,dose==.5)
t.test(y.cmp$len.vc,y.cmp$len.oj,paired=TRUE)$conf;

## [1] -9.236542 -1.263458
## attr(,"conf.level")
## [1] 0.95

t.test(y.cmp$len.vc,y.cmp$len.oj,paired=TRUE)$p.value;

## [1] 0.01547205
```

We can see the p-value is less than 5% and confidence level doesn't include 0. Then we can reject  $H_0$  and accept  $H_a$ . Similarly we can test dose levels of 1&2. Dose Level = 1

```
## [1] -9.908089 -1.951911
## attr(,"conf.level")
## [1] 0.95

## [1] 0.008229248
```

Dose Level = 2

```
## [1] -4.168976 4.328976
## attr(,"conf.level")
## [1] 0.95

## [1] 0.9669567
```

Now, we can see difference in means under dose level of 0.5 and 1 are significantly not equal to 0, but not at the level of 2. (p.value=0.97)

Our conclusion is that supp and dose are both dominant factors to length of tooth-growth, but the dominance is not significant under few situations.