

ToothGrowth Data Analyses

By Ling Tian, Feb.2015

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
data(ToothGrowth)
head(ToothGrowth)
```

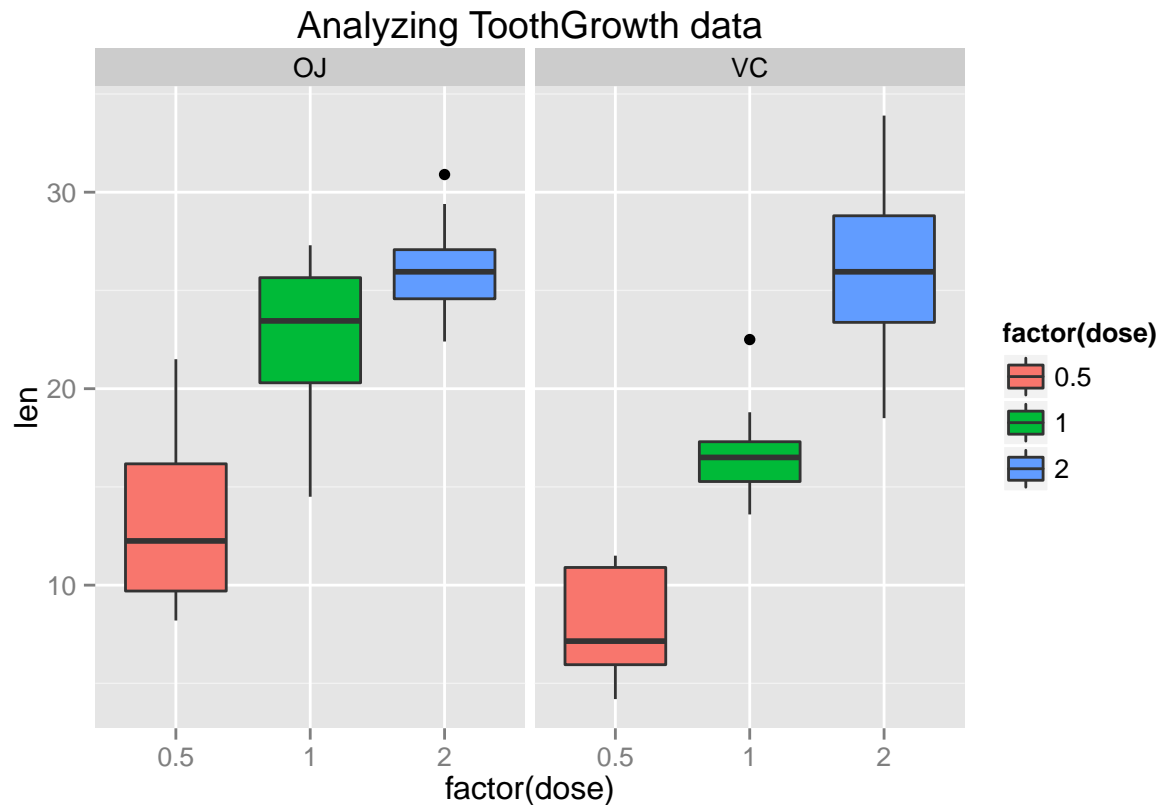
```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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 ...
```

Perform box-plot on data

```
library(ggplot2)
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose)))+geom_boxplot()+facet_grid(.~supp)+gg
```



OJ and VC, as the dose increases, the tooth length grows.

Basic summary of the data

```
summary(ToothGrowth)
```

```
##      len      supp  dose
##  Min.   : 4.20   OJ:30  0.5:20
##  1st Qu.:13.07   VC:30  1 :20
##  Median :19.25           2 :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

There are two treatment(OJ and VC),each treatment has three dose(0.5, 1, and 2)

Compare tooth growth by supp and dose

Calculate the confidence intervals

```
mean_VC<-mean(ToothGrowth$len[1:30])
mean_OJ<-mean(ToothGrowth$len[31:60])
var_VC<-(sd(ToothGrowth$len[1:30]))^2
var_OJ<-(sd(ToothGrowth$len[31:60]))^2
```

```
q<-(((var_VC+var_OJ)/30)^2)/(((var_VC/30)^2+((var_OJ/30)^2))/29)
t<-qt(0.975,q)##for 95% confidence intervals,2.5% on each side
mean_OJ-mean_VC+c(-1,1)*t*sqrt(var_VC/30+var_OJ/30)
```

```
## [1] -0.1710156  7.5710156
```

Hypothesis test

H0: Supplement types have no effect on tooth length

```
t.test(len~supp,data=ToothGrowth,paired=FALSE)
```

```
##
## 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
```

The confidence interval contains 0, so the difference between the two treatment is not significant. The p-value=0.06, which is bigger than 0.05, so we fail to reject the hypothesis.

H0: Supplement types and dose have no effect on tooth length.

```
Low<-subset(ToothGrowth,dose==0.5)
t.test(len~supp,data=Low,paired=FALSE)
Med<-subset(ToothGrowth,dose==1)
t.test(len~supp,data=Med,paired=FALSE)
High<-subset(ToothGrowth,dose==2)
t.test(len~supp,data=High,paired=FALSE)
```

Analysis of two treatments with low/median/high dose on tooth length: For low dose, $P=0.006$ which is smaller than 0.05, so different treatments do have an effect on tooth length when dose is low. For median dose, $P=0.001$ which is smaller than 0.05, so different treatments do have an effect on tooth length when dose is median. For high dose, $P=0.96$ which is much bigger than 0.05, so we fail to reject the hypothesis, and different treatments do not have an effect on tooth length when dose is high.

Conclusion

At low dosage ($=0.5$ or 1), OJ treatment yield a greater tooth growth than VC treatment. However, when the dosage is at a higher level ($=2$), there shows no significant difference between the two treatments on tooth growth.

Appendix

Analysis results of two treatments with low/median/high dose on tooth length

```
Low<-subset(ToothGrowth,dose==0.5)
t.test(len~supp,data=Low,paired=FALSE)
```

```
##
## 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
```

```
Med<-subset(ToothGrowth,dose==1)
t.test(len~supp,data=Med,paired=FALSE)
```

```
##
## 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
```

```
High<-subset(ToothGrowth,dose==2)
t.test(len~supp,data=High,paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = -0.0461, 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
```