Statistical Inference Course Project - Part 2

Li Xin
23 December, 2015

Overview

This is the second part of project for the statistical inference class. In this part, ToothGrowth dataset will be loaded and performed some basic exploratory analysis.

Load Dataset

```
library(datasets)
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

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

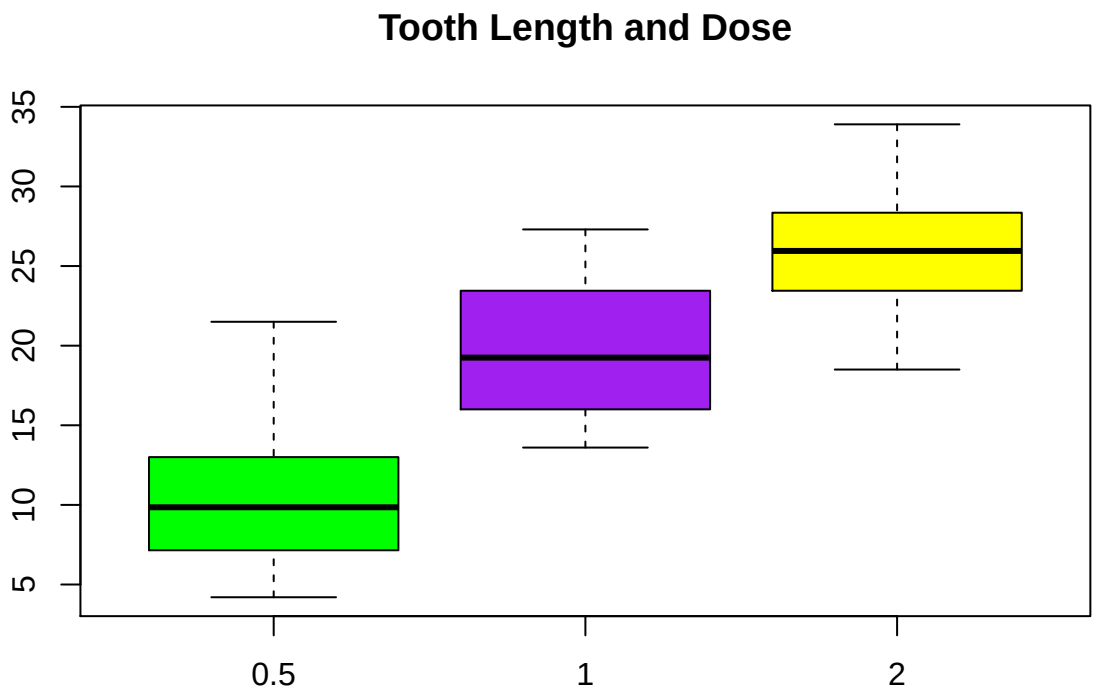
Basic Summary

```
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
```

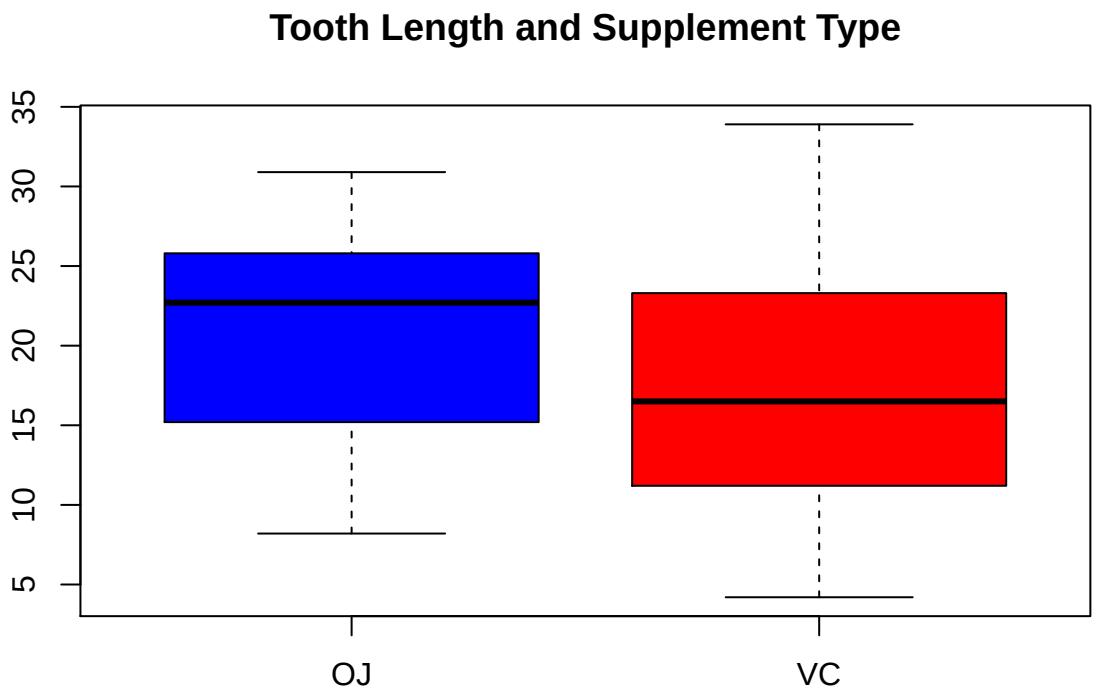
The below box plot shows tooth length with Dose. We can see there is a big difference there.

```
boxplot(len~dose,data=ToothGrowth,col=c("green","purple","yellow"),
        main="Tooth Length and Dose")
```



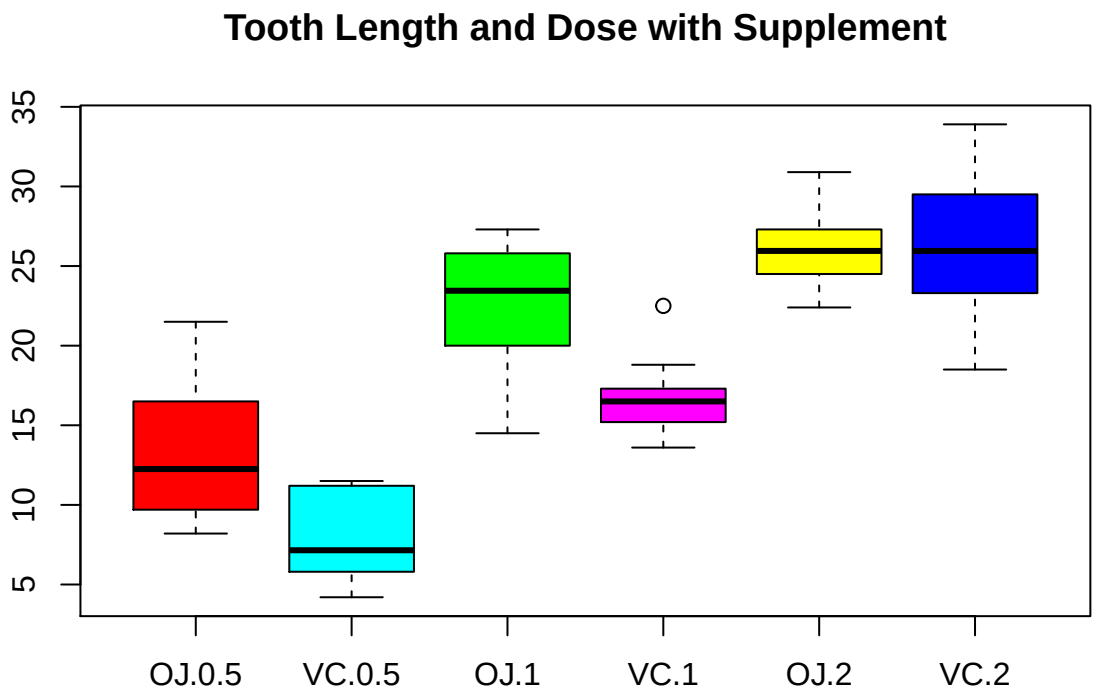
The below box plot shows tooth length with supplement type. There are not much difference between these 2 types of supplement.

```
boxplot(len~supp,data=ToothGrowth,col=c("blue","red"),
        main="Tooth Length and Supplement Type")
```



The below box plot shows interaction between dose and supplement type.

```
boxplot(len~interaction(supp,dose),data=ToothGrowth,
        main="Tooth Length and Dose with Supplement",col=sample(rainbow(6)))
```



From the above box plot we can see supplement types increases with the tooth length.

Confidence Interval & Hypothesis

From the above plots we can assume that for tooth length, there is probably a highly significant difference between dose, and not a significant difference between supplement type.

Below is the T test for supplement type. Then test the effect of supplement type on tooth length when have dose constant at each level.

```
oj <- ToothGrowth[ToothGrowth$supp=="OJ",]
vc <- ToothGrowth[ToothGrowth$supp=="VC",]
oj05 <- oj[oj$dose==0.5,]
oj10 <- oj[oj$dose==1.0,]
oj20 <- oj[oj$dose==2.0,]
vc05 <- vc[vc$dose==0.5,]
vc10 <- vc[vc$dose==1.0,]
vc20 <- vc[vc$dose==2.0,]

#t test vc and oj then dose = 0.5
vcoj0.5 <- t.test(len~supp,data=rbind(vc05,oj05),var.equal=FALSE)

#t test vc and oj then dose = 1.0
vcoj1.0 <- t.test(len~supp,data=rbind(vc10,oj10),var.equal=F)
```

```r
#t test vc and oj then dose = 2.0
vcoj2.0 <- t.test(len~supp,data=rbind(vc20,oj20),var.equal=F)
```

Here we can see the the summary:

For vc and oj at dose=0.5: the P value is 0.0063586, the conf int (-) is 1.7190573, the conf int (+) is 8.7809427

For vc and oj at dose=1.0: the P value is 0.0010384, the conf int (-) is 2.8021482, the conf int (+) is 9.0578518

For vc and oj at dose=2.0: the P value is 0.9638516, the conf int (-) is -3.7980705, the conf int (+) is 3.6380705

Here we are going to test the effect of dose on tooth length:

```r
#0.5 vs 1.0
vc0.5and1.0 <- t.test(len~dose,data=rbind(vc05,vc10),var.equal=TRUE)
oj0.5and1.0 <- t.test(len~dose,data=rbind(oj05,oj10),var.equal=TRUE)

#1.0 vs 2.0
vc1.0and2.0 <- t.test(len~dose,data=rbind(vc10,vc20),var.equal=TRUE)
oj1.0and2.0 <- t.test(len~dose,data=rbind(oj10,oj20),var.equal=TRUE)

#0.5 vs 2.0
vc0.5and2.0 <- t.test(len~dose,data=rbind(vc05,vc20),var.equal=TRUE)
oj0.5and2.0 <- t.test(len~dose,data=rbind(oj05,oj20),var.equal=TRUE)
```

Here we can see the summary:

For VC 0.5 and 1.0: P-value is 6.4922646×10^{-7}, conf int (-) is -11.2643455, conf int (+) is -6.3156545

For OJ 0.5 and 1.0: P-value is 8.3575593×10^{-5}, conf int (-) is -13.4108143, conf int (+) is -5.5291857

For VC 1.0 and 2.0: P-value is 3.3975779×10^{-5}, conf int (-) is -12.9689598, conf int (+) is -5.7710402

For OJ 1.0 and 2.0: P-value is 0.0373628, conf int (-) is -6.5005017, conf int (+) is -0.2194983

For VC 0.5 and 2.0: P-value is 4.9572857×10^{-9}, conf int (-) is -21.8328433, conf int (+) is -14.4871567

For OJ 0.5 and 2.0: P-value is 3.4018585×10^{-7}, conf int (-) is -16.2782226, conf int (+) is -9.3817774

Conclusion

From the analysis above, we can conclude there is a definite dependence on the dose level on tooth growth. P-value are less than threhold and confidence intervals don't include zero. This indicate that the when supplement dose increase, the tooth growth will increase significantly.