ToothGrowth Data Analysis

liwenlong

Overview

In this report, I will perform some basic explortary data analyses on the ToothGrowth data, And compare tooth growth by supp and dose.

```r
library(dplyr)
library(datasets)
library(ggplot2)
```

Read data

- View the structure of the data

```r
#convert the dose from num to factor
ToothGrowth$dose <-as.factor(ToothGrowth$dose)
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 ...
```

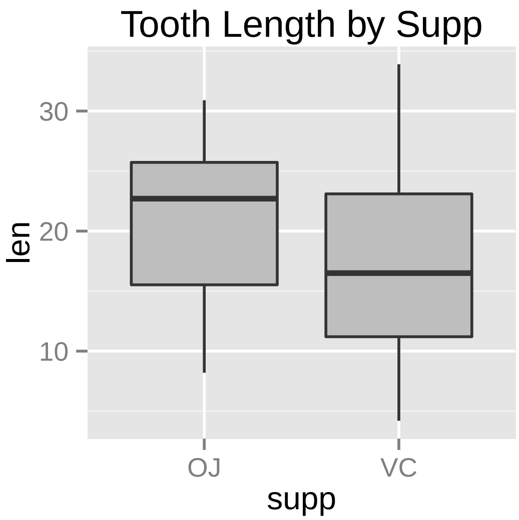
```r
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##     
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

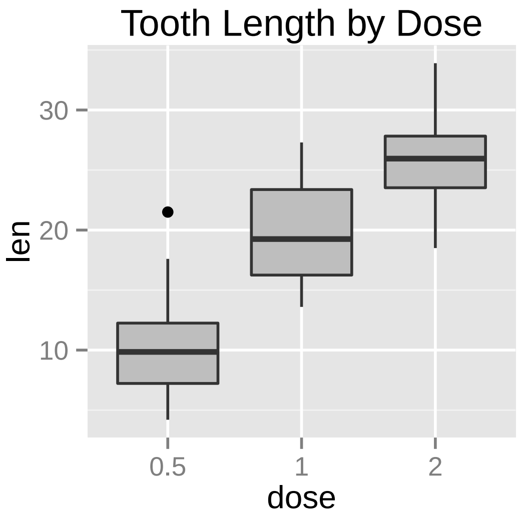
We have sixty records group by supp(QJ,VC) and dose (0.5,1,2), and for each group there is 10 records.

- Explor the data by dose and supp

```r
#summary by supp
ggplot(ToothGrowth,aes(x=supp,y=len))+
        geom_boxplot(fill="grey")+
        ggtitle("Tooth Length by Supp")
```



```
#summary by dose
ggplot(ToothGrowth,aes(x=dose,y=len))+
        geom_boxplot(fill="grey")+
        ggtitle("Tooth Length by Dose")
```



Tooth length compare by Supp

- Assumption is there is no difference between two Supp group(They have the same mean value).
- I didn't read the description of this dataset, but from the definition of the problem, it doesn't tell us the data is paired or they have the same variance. So i will set {paired = F,var.equal = F}

```
testBySupp <- t.test(subset(ToothGrowth$len,ToothGrowth$supp=="OJ"),
       subset(ToothGrowth$len,ToothGrowth$supp=="VC"),
       var.equal = F,
       paired = F)
c(testBySupp$conf,testBySupp$p.value)
```

```
## [1] -0.17101562  7.57101562  0.06063451
```

- The 95 percent confdence interval[**-0.1710156, 7.5710156**] contains 0. And the P-Value [**0.0606345**] is greater than 0.05.
- So we do not have conclusive evidence to show that OJ has a better effect on tooth growth and failed to reject the null hypothesis.
- Actually I also tried other params for the test,setting {var.equal=T} doesn't make any difference, while setting {paired = T} will lead to a totally different result).

Tooth length compare by Dose

- The Null Hypothesis is the mean value are the same for different dose group.
- User {paired = F,var.equal = F} for the test, same reason as above.

```
test1<-t.test(subset(ToothGrowth$len,ToothGrowth$dose==1),
       subset(ToothGrowth$len,ToothGrowth$dose==.5),
       var.equal = F,paired = T)
test2<-t.test(subset(ToothGrowth$len,ToothGrowth$dose==2),
       subset(ToothGrowth$len,ToothGrowth$dose==1),
       var.equal = F,paired = T)
c(test1$conf,test1$p.value)
```

```
## [1] 6.387121e+00 1.187288e+01 1.225437e-06
```

```
c(test2$conf,test2$p.value)
```

```
## [1] 3.4718143442 9.2581856558 0.0001934186
```

- From both tests, the 95-percent-interval is above 0 with p-value < .05 which provides statistical evidence that higher Dose DOES have positive impact on tooth length.
- Setting [var.equal = T] shows the same result, while setting {paired = T} leads to a slightly different result, but the conclusion is still the same.