Statistical Inference Course Project - Part II

In this part, we will do some analyzing to the ToothGrowth data in the R datasets package

Loading the ToothGrowth data and perform some basic exploratory data analyses

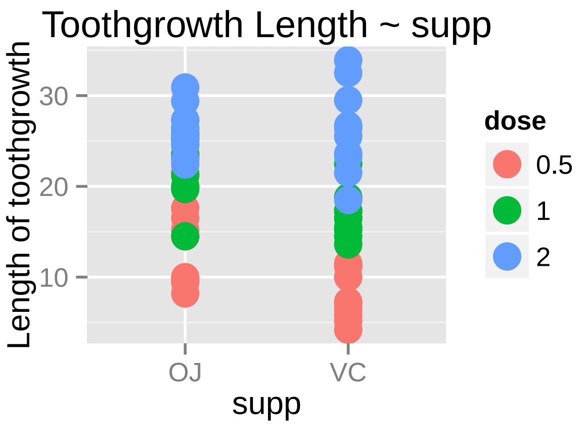
The data loading is with the command: data("ToothGrowth"). Then,we can do some exploratory for the ToothGrowth data set:

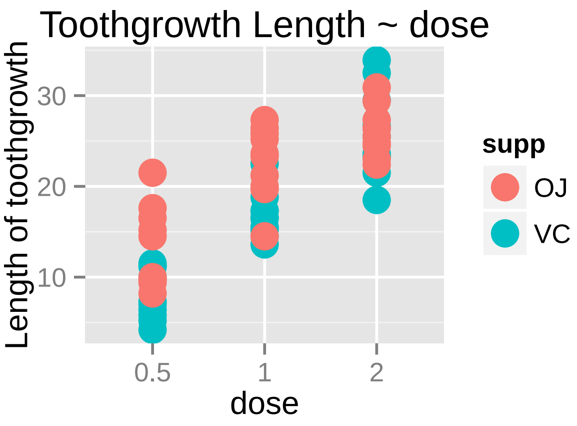
With str() function, we can get the data fields names, types and some initial vaues as:

```
## '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 ...
```

From here, we know there are three variables "len","supp" and "dose" and total 60 bservation in the data set.

Now, we can show these data with graphically expresion to make more clear:





From above graph, we can see that toothgrow length can be grouped by dose and supp variables.

Provide a basic summary of the data

Now, we need check ToothGrowth's more detail statistics info. Let's start from summary(ToothGrowth):

```
##       len        supp     dose   
##  Min.   : 4.2   OJ:30   0.5:20  
##  1st Qu.:13.1   VC:30   1  :20  
##  Median :19.2           2  :20  
##  Mean   :18.8                   
##  3rd Qu.:25.3                   
##  Max.   :33.9                   
```

This gives us the total mean length as 18.8133. We also can get length standard deviation as 7.6493.

We also can get each group's mean lenght and standard deviation as:

Grouped by 'supp': means are 20.6633, 16.9633, sds are 6.6056, 8.266.

Grouped by 'dose': means are 10.605, 19.735, 26.1, sds are 4.4998, 4.4154, 3.7742.

Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

From previous discussion, we can find that there are equal elements in each supp or dose group. But there is no information can indicate whether the testing is paired or non-paired. So the hypothesis testing will be performed in paired, non-paired, and non-paried and equal-variance these three assumptions.

The confidence intercal results grouped by supp is:

```
##                       CIlo  CIhi
## paired               1.409 5.991
## unpaired            -0.171 7.571
## unpaired(equalVar) -0.167 7.567
```

From this ouptput, we can say that there is a confidence difference between OJ and VC if the paired hypoyhesis is true. For both of the unpaired testing contains the zero, we cannot say there is a confidence difference between OJ and VC if unpaired assumption is true.

The comparing results between different dose values are:

```
##                    CIl(0.5~1.0) CIh(0.5~1.0 CIl(1.0~2.0) CIh(1.0~2.0)
## paired                    6.387       11.87        3.472        9.258
## unpaired                  6.276       11.98        3.734        8.996
## unpaired(equalVar)        6.276       11.98        3.736        8.994
##                    CIl(0.5~2.0) CIh(0.5~2.0)
## paired                    12.62        18.37
## unpaired                  12.83        18.16
## unpaired(equalVar)        12.84        18.15
```

From the testing results, we can find that any two different groups based on dose's three confidence interval testings are above zero. This shows the bigger dose intaking is better for the tooth growth.

Conclusions and the assumptions

From above data exploratory and hypothesis testing, we can have the following conclusions:

- Each dose or supp based group has the same number of the elements
- There's no other information indicating the different groups are paired or non-paired.
- So the hypothesis testing will be performed on paired, non-paired, and non-paried and equal-variance
- The supp based groups testing shows there isn't confidence difference between VC and OJ on unpaired assumptions
- The supp based groups testing shows there is confidence difference between VC adn OJ on paired assumptions
- Different dose group testing show there are confidence difference between groups on all three assumptions
- Different dose intaking is more efficitvie for tooth growth than differnt supp
- The bigger dose taking, the better for tooth growth

Appendix

This appendix is used to show the simulation codes used in the report.

```
data("ToothGrowth")
str(ToothGrowth)

ToothGrowth$dose=as.factor(ToothGrowth$dose)
ToothGrowth$supp=as.factor(ToothGrowth$supp)

library(ggplot2)
g <- ggplot(ToothGrowth, aes(x = supp, y = len, group = dose))
g <- g + geom_point(size = 5, aes(colour = dose))
g + labs(x = "supp", y = "Length of toothgrowth")+ggtitle("Toothgrowth Length ~ supp")
```

```
g <- ggplot(ToothGrowth, aes(x = dose, y = len, group = supp))
g <- g + geom_point(size = 5, aes(colour = supp))
g + labs(x = "dose", y = "Length of toothgrowth")+ggtitle("Toothgrowth Length ~ dose")
```

```
summary(ToothGrowth)
sd(ToothGrowth$len)

with(ToothGrowth,tapply(len,supp,mean))
with(ToothGrowth,tapply(len,supp,sd))

with(ToothGrowth,tapply(len,dose,mean))
with(ToothGrowth,tapply(len,dose,sd))

groupsComare<-function(grp1,grp2){
  g1 <- grp1$len; g2 <- grp2$len;

  rbind(
    as.vector(t.test(g2,g1,paired = TRUE)$conf.int),
    as.vector(t.test(g2, g1)$conf.int),
    as.vector(t.test(g2,g1,var.equal=TRUE)$conf.int)
  )
}
## slicing the original data by supp and dose
supp.vc <- subset(ToothGrowth,supp=="VC")
supp.oj <- subset(ToothGrowth,supp=="OJ")

dose.0.5<- subset(ToothGrowth,dose==0.5)
dose.1.0<- subset(ToothGrowth,dose==1.0)
dose.2.0<- subset(ToothGrowth,dose==2.0)

rnames<-c("paired","unpaired","unpaired(equalVar)")

suppnames<-c("CIlo","CIhi")
dosenames<-c("CIl(0.5~1.0)","CIh(0.5~1.0","CIl(1.0~2.0)",
             "CIh(1.0~2.0)","CIl(0.5~2.0)","CIh(0.5~2.0)")
## Compare tooth growth by supp
## Null hypothesis: True difference in means is equal to 0
supp<-groupsComare(supp.vc,supp.oj)
```

```r
rownames(supp)<-rnames
colnames(supp)<-suppnames
supp

dose<-cbind(groupsComare(dose.0.5,dose.1.0),groupsComare(dose.1.0,dose.2.0),
            groupsComare(dose.0.5,dose.2.0))
rownames(dose)<-rnames
colnames(dose)<-dosenames
dose
```