

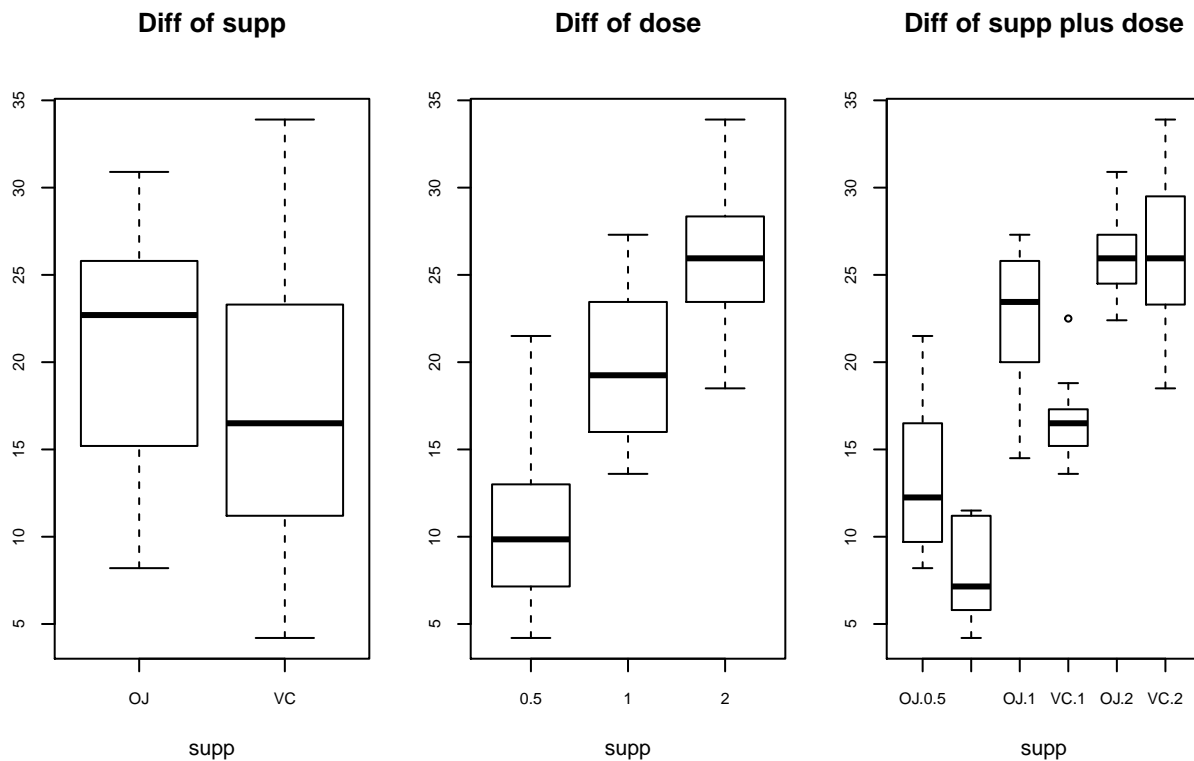
Toothgrowth data analysis

Yilin

February 12, 2016

1. Load the ToothGrowth data and perform basic exploratory data analyses

```
library(datasets)
data(ToothGrowth)
par(mfrow=c(1,3),mar=c(6,2,6,2))
boxplot(ToothGrowth$len~I(ToothGrowth$supp),xlab="supp", ylab="length",
        cex.axis=0.8,main="Diff of supp")
boxplot(ToothGrowth$len~I(ToothGrowth$dose),xlab="supp", ylab="length",
        cex.axis=0.8,main="Diff of dose")
boxplot(ToothGrowth$len~I(ToothGrowth$supp)+I(ToothGrowth$dose),
        xlab="supp", ylab="length",cex.axis=0.8,main="Diff of supp plus dose")
```



2. Basic summary of the data

```
summary(ToothGrowth)
```

```
##      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
```

```
with(ToothGrowth, table(supp, dose))
```

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

```
mean_len<-aggregate(len ~ ., data=ToothGrowth, mean)
mean_len
```

```
##   supp dose   len
## 1    OJ  0.5 13.23
## 2    VC  0.5  7.98
## 3    OJ  1.0 22.70
## 4    VC  1.0 16.77
## 5    OJ  2.0 26.06
## 6    VC  2.0 26.14
```

3. Compare Tooth growth by supp and dose

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
D0.5<-subset(ToothGrowth, dose==0.5); D1<-subset(ToothGrowth, dose==1)
D2<-subset(ToothGrowth, dose==2); OJ<-subset(ToothGrowth, supp=="OJ")
VC<-subset(ToothGrowth, supp=="VC")
D0.5_1=t.test(D0.5$len, D1$len)$conf.int
D1_2=t.test(D1$len, D2$len)$conf.int
D0.5_2=t.test(D0.5$len, D2$len)$conf.int
OJ_VC=t.test(OJ$len, VC$len)$conf.int
conf.int<-cbind(group=c("D0.5_1", "D1_2", "D0.5_2", "OJ_VC"),
                 low_diff=c(D0.5_1[1], D1_2[1], D0.5_2[1], OJ_VC[1]),
                 high_diff=c(D0.5_1[2], D1_2[2], D0.5_2[2], OJ_VC[2]))
conf.int
```

```
##      group      low_diff      high_diff
## [1,] "D0.5_1" "-11.9837812579016" "-6.27621874209841"
## [2,] "D1_2"   "-1.87410125817549"  "3.71743459150882"
## [3,] "D0.5_2" "-11.0321173579622"  "-5.38454930870448"
## [4,] "OJ_VC"  "-0.171015618367165"  "7.57101561836716"
```

4.State conclusions and assumptions Conclusion: 1.Different methods of intaking the drug between VC and OJ do impact toothgrowth, the toothlength of OJ group is longer than VC group. 2. Increaseing the dose of the drug from low(0.5) to median(1) significantly promotes toothgrowth. However, further increasing drug dosage from 1 to 2 does not accelerate toothgrowth. Assumption: We define the type I error rate at $\alpha=0.05$, that means the false rate of rejecting the H_0 (toothlength between the compared two group is equal) when the H_0 is actually true is 5%.