Part 2: Basic Inferential Data Analysis Instructions

Chuwei Zhong

October 5, 2019

Load data && simple analysis

```r
library(ggplot2)
```

```r
data(ToothGrowth)

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 ...
```

```r
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
```

```r
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  
```

```r
dim(ToothGrowth)
```

```
## [1] 60  3
```

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

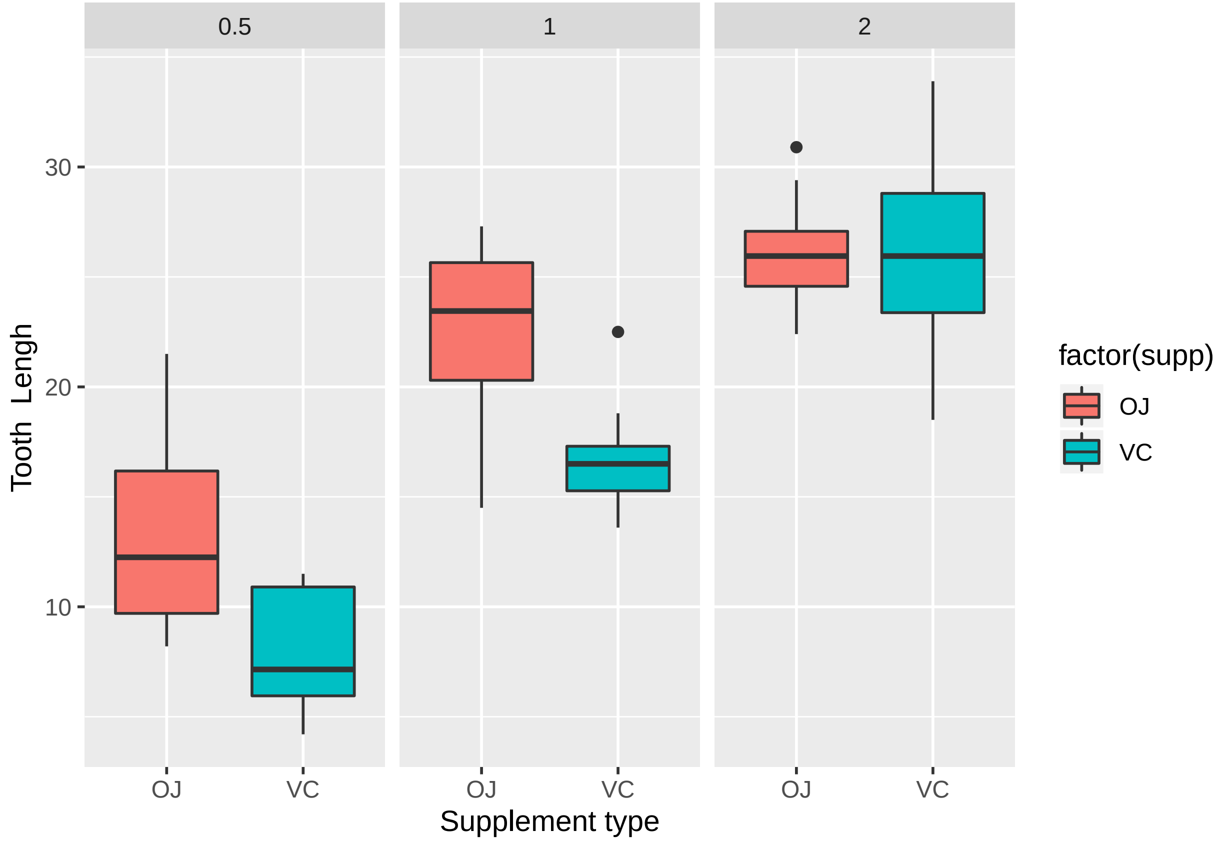
```
## 
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

basic summary

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(ToothGrowth, aes(factor(supp), len, fill = factor(supp))) +
      geom_boxplot() + facet_grid(.~dose) +xlab("Supplement type") + ylab("Tooth  Lengh")
```



Use confidence intervals and/or hypothesis tests to compare tooth growth by supply and dose

Here,null hypothesis is: tooth growth has nothing to do with the supply in different levels of dosages.

```
#compare the influence by delivery method for the same dosage
data05 <- rbind(ToothGrowth[(ToothGrowth$dose == 0.5) & (ToothGrowth$supp == "OJ"),], ToothGrowth[(Toot
t05 <- t.test(len ~ supp, data05, var.equal = FALSE, conf.level = 0.95)

data1 <- rbind(ToothGrowth[(ToothGrowth$dose == 1) & (ToothGrowth$supp == "OJ"),], ToothGrowth[(ToothGr
t1 <- t.test(len ~ supp, data1, var.equal = FALSE, conf.level = 0.95)

data2 <- rbind(ToothGrowth[(ToothGrowth$dose == 2) & (ToothGrowth$supp == "OJ"),], ToothGrowth[(ToothGr
```

```r
t2 <- t.test(len ~ supp, data2, var.equal = FALSE, conf.level = 0.95)

#extract p-values and confidence intervals
summary_result <- data.frame(
  c(t05$p.value, t1$p.value, t2$p.value),
  c(t05$conf.int[1],t1$conf.int[1], t2$conf.int[1]),
  c(t05$conf.int[2],t1$conf.int[2], t2$conf.int[2])
)
row.names(summary_result) <- c("Dosage 0.5","Dosage 1.0","Dosage 2.0")
colnames(summary_result) <- c("p.value","Conf.Low","Conf.High")
summary_result
```

```
##                 p.value  Conf.Low Conf.High
## Dosage 0.5 0.006358607  1.719057  8.780943
## Dosage 1.0 0.001038376  2.802148  9.057852
## Dosage 2.0 0.963851589 -3.798070  3.638070
```

Here, for dosages to be 0.5mg and 1.0mg, p-values are less than 0.05 and confidence intervals don't include 0. So we reject the null hypothesis. For dosage to be 2.0mg, p-values are larger than 0.05 and confidence intervals include 0. Hence, we fail to reject the null hypothesis.

State my conclusions and assumptions

In lower dosages(0.5-1.0mg), orange juice can improve tooth growth than Vitamin C. While in higher dosage(2.0mg), tooth growth rate is not statistically different between these two supplyment methods.