Statistical Reference Course Assignment, Part Two

Xiaocheng Zeng

January 19, 2016

Overview

This analysis attempts to provide some insights of ToothGrow data. We investigate the data with exploratory analysis and hypothesis tests/confidence interval for different subsets of data.

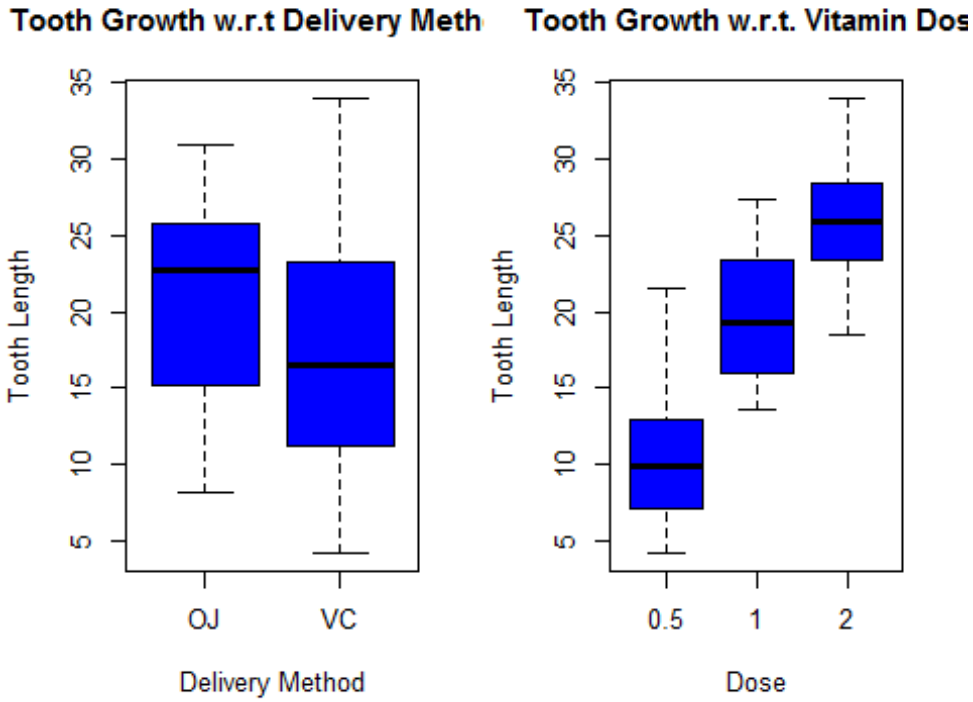
Exploratory analysis

We provide a general summary statistics of the whole sample. According to sample description, the subjects can be divided into different groups based on different Vitamin delivery methods and Vitamin dose. We create boxplot for tooth length w.r.t the two factor variables.

```
data(ToothGrowth)
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
```

```
par(mfrow=c(1,2),cex=.8)
boxplot(len~supp,data=ToothGrowth,main="Tooth Growth w.r.t Delivery
Methods",xlab="Delivery Method",ylab="Tooth Length",col="blue")
boxplot(len~dose,data=ToothGrowth,main="Tooth Growth w.r.t. Vitamin
Dose",xlab="Dose",ylab="Tooth Length",col="blue")
```



From the boxplots, we make the following 3 hypotheses,

1. Teeths are longer for subjects receiving orange juice (OJ) than those receiving ascorbic acid (VC)
2. Teeths are longer for subjects receiving heavier dose of Vitamin. Therefore, subjects receiving dose=2 have longer teeth than those receiving dose=1,
3. while subject receiving dose=.5 have shorter teeth than those receiving dose=1.

Hypothesis tests based on subset data

```
supp_oj<-subset(ToothGrowth,supp=="OJ")
supp_vc<-subset(ToothGrowth,supp=="VC")
dose_.5<-subset(ToothGrowth,dose==.5)
dose_1<-subset(ToothGrowth,dose==1)
dose_2<-subset(ToothGrowth,dose==2)
```

test of 1st hypothesis, H_0: mu(len)_oj-mu(len)_vc=0 against H_a: mu(len)_oj-mu(len)_vc>0

```
t.test(supp_oj$len-supp_vc$len,paired=FALSE,var.equal=FALSE)$p.value

## [1] 0.002549842

t.test(supp_oj$len-supp_vc$len,paired=FALSE,var.equal=FALSE)$conf
```

```
## [1] 1.408659 5.991341
## attr(,"conf.level")
## [1] 0.95
```

test of 2nd hypothesis, H_0:mu(len)_dose2-mu(len)_dose1=0 against H_a:mu(len)_dose2-mu(len)_dose1>0

```
t.test(dose_2$len-dose_1$len,paired=FALSE,var.equal=FALSE)$p.value
```

```
## [1] 0.0001934186
```

```
t.test(dose_2$len-dose_1$len,paired=FALSE,var.equal=FALSE)$conf
```

```
## [1] 3.471814 9.258186
## attr(,"conf.level")
## [1] 0.95
```

test of 3nd hypothesis, H_0:mu(len)_dose.5-mu(len)_dose1=0 against H_a: mu(len)_dose.5-mu(len)_dose1<0

```
t.test(dose_.5$len-dose_1$len,paired=FALSE,var.equal=FALSE)$p.value
```

```
## [1] 1.225437e-06
```

```
t.test(dose_.5$len-dose_1$len,paired=FALSE,var.equal=FALSE)$conf
```

```
## [1] -11.872879  -6.387121
## attr(,"conf.level")
## [1] 0.95
```

Conclusions and assumptions for conclusions

Conclusions

1.p value for null hypothsis 1 is far smaller than .05 and H_0 is rejected. We conclude Teeths are longer for subjects receiving orange juice (OJ) than those receiving ascorbic acid (VC).

2.p value for null hypotheses 2 and 3 are far smaller than .05 and both are rejcted. We conclude subjects receiving dose=2 have longer teeth than those receiving dose=1 while subject receiving dose=.5 have shorter teeth than those receiving dose=1. Therefore, subjects receiving heavier dose have longer teeth in general.

Assumptions

1.The subjects receiving different delivery methods and doses are independent. In another word, they are not paired observations.

2.The subjects with different delivery methods belong to two different populations. And therefore we assume that they have different variances. We have the same assumptions for subjects receiving different doses.