Tooth Growth rate by varying source of vitamin intake

Cy

11/4/2020

Part 2

Overview

In this section, Exploratory data analysis will be performed on Tooth Growth data. The performance of using Orange Juice and Vitamin C with different dosages are evaluated through statistical inference methodology.

```
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
```

```
head(ToothGrowth,3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

Boxplot for tooth growth

```
data("ToothGrowth")
ToothGrowth$supp<-as.factor(ToothGrowth$supp)
boxplot(len~dose+supp, data=ToothGrowth,col='orange')
```

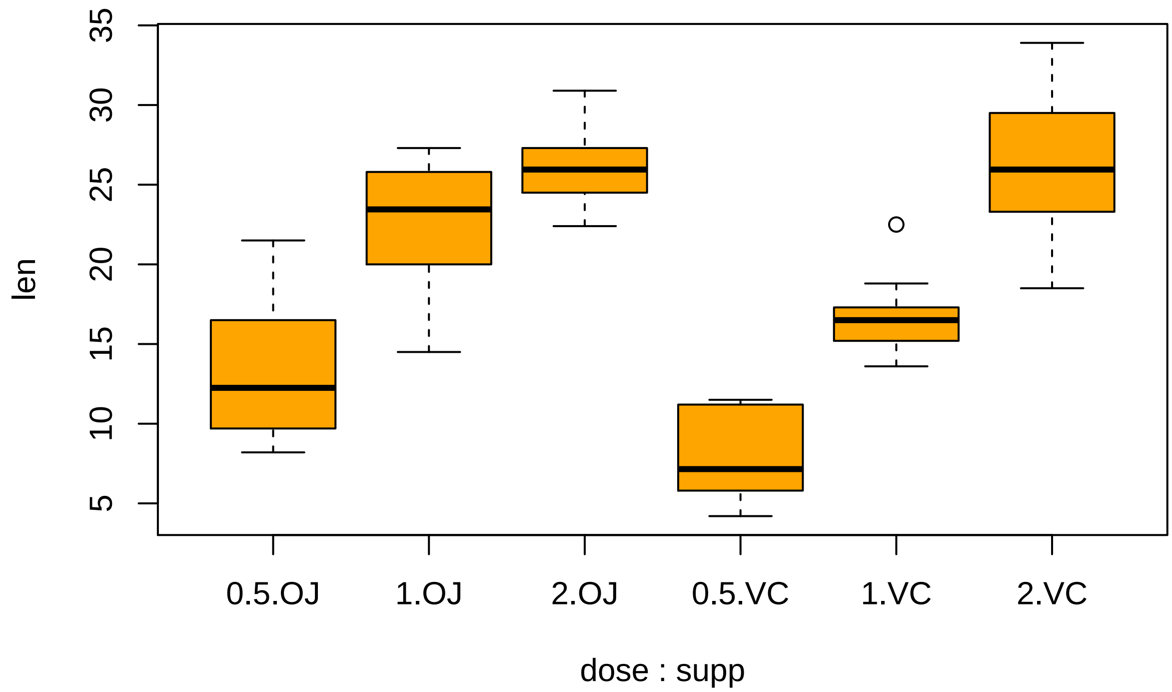

Figure above visualize the Tooth Growth Data, Through visual inspection, The Orange Juice has better performance

Hypothesis testing and Confidence Interval

Using T-distribution since each unique case has less samples.
Null hypothesis : Orange Juice has same effect with Vitamin C
Alternative hypothesis : Orange Juice has better effect compared to Vitamin C
To obtain better data consistency, the dosage was varied and compared to each other

```
# 0.5 Dosage
oj5<-subset(ToothGrowth,supp=="OJ" & dose=="0.5",select =len)
vc5<-subset(ToothGrowth,supp=="VC"& dose=="0.5",select =len)
p5<-t.test(oj5,vc5,paired = FALSE)

# 1 Dosage
oj1<-subset(ToothGrowth,supp=="OJ" & dose=="1",select =len)
vc1<-subset(ToothGrowth,supp=="VC"& dose=="1",select =len)
p1<-t.test(oj1,vc1,paired = FALSE)

# 2 Dosage
oj2<-subset(ToothGrowth,supp=="OJ" & dose=="2",select =len)
vc2<-subset(ToothGrowth,supp=="VC"& dose=="2",select =len)
p2<-t.test(oj2,vc2,paired = FALSE)

result<-data.frame("Dosage 0.5"=c(p5$p.value,p5$conf.int[1],p5$conf.int[2]),
```

```r
           "Dosage 1.0"=c(p1$p.value,p1$conf.int[1],p1$conf.int[2]),
           "Dosage 2.0"=c(p2$p.value,p2$conf.int[1],p2$conf.int[2]),
           row.names = c("p-value","Lower Confidence Limit","Upper Confidence Limit"))
print(round(result,4))
```

```
##                        Dosage.0.5 Dosage.1.0 Dosage.2.0
## p-value                    0.0064     0.0010     0.9639
## Lower Confidence Limit     1.7191     2.8021    -3.7981
## Upper Confidence Limit     8.7809     9.0579     3.6381
```

Conclusion (Part II)

The p-values for 0.5 and 1.0 dosage are below the threshold 0.05, proving the improvement of using **orange juice** compared to **vitamin C**. However, the p-value for 2.0 dosage is above 0.05 and its lower confidence limit is **below zero**. This may arisen from faulty data or poor methodology of research. Considering cases for 2.0 dosage, null hypothesis cannot be rejected.

Recommendation

The experiment should be conducted with more data samples to obtain a much convincing inference with solid evidence.