Final Project. Part 2: Basic Inferential Data Analysis Instructions

Kadyrbek

11/18/2019

Synopsis

In this project we are going to analyze ToothGrowth data from datasets package. We will compare the length of odontoblasts by supplement types and doses amounts.

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

Here is some exploratory data analyses of ToothGrowth data

```
library(datasets)
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
```

```
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 ...
```

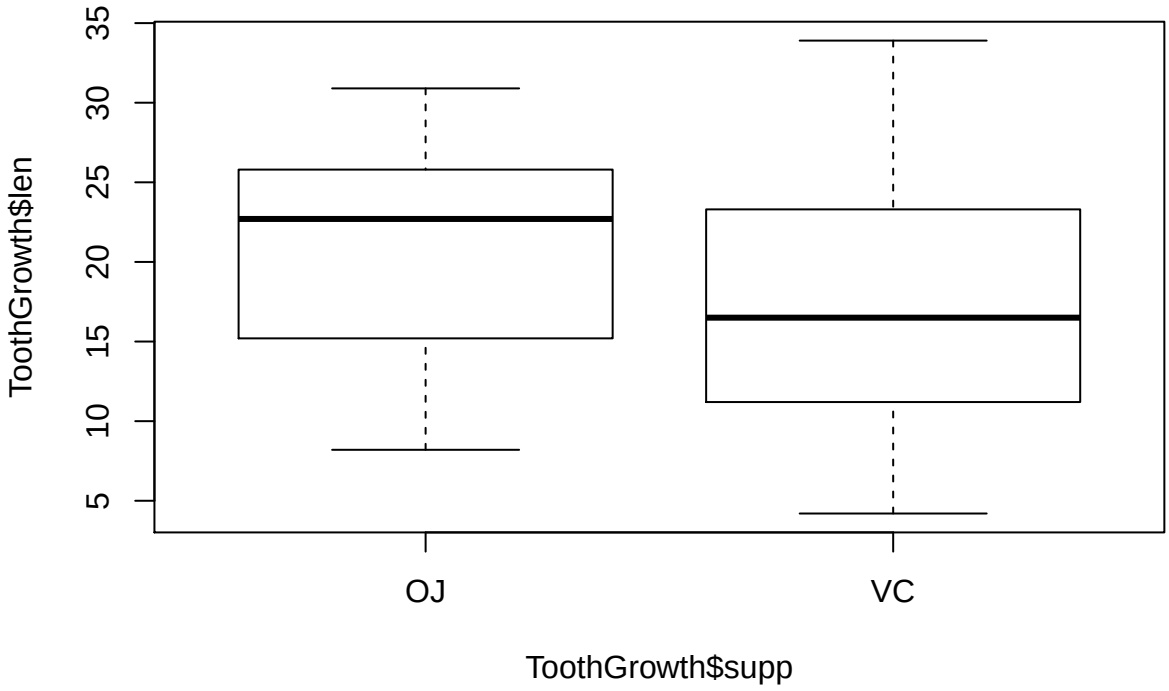
#2. Provide a basic summary of the data.

```
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
```

From the table above we can see that there are 60 observations with two supplement types 30 observation each. There are 3 types of dose amounts 0.5, 1 and 2.

```
boxplot(ToothGrowth$len~ToothGrowth$supp)
```



#3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Here are confidence intervals for supplement types.

```
oj<-ToothGrowth[ToothGrowth$supp=="OJ",]
vc<-ToothGrowth[ToothGrowth$supp=="VC",]
ojt<-t.test(oj$len, alternative = "two.sided",conf.level = .8)$conf
vct<-t.test(vc$len, alternative = "two.sided",conf.level = .8)$conf
list("OJ interval"=ojt,"VC interval"=vct)
```

```
## $`OJ interval`
## [1] 19.08174 22.24493
## attr(,"conf.level")
## [1] 0.8
##
## $`VC interval`
## [1] 14.98417 18.94250
## attr(,"conf.level")
## [1] 0.8
```

```
t.test(oj$len,vc$len)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  oj$len and vc$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Here are confidence intervals for different dose amounts.

```
half<-ToothGrowth[ToothGrowth$dose==0.5,]
one<-ToothGrowth[ToothGrowth$dose==1,]
two<-ToothGrowth[ToothGrowth$dose==2,]
ht<-t.test(half$len, alternative = "two.sided")$conf
ot<-t.test(one$len, alternative = "two.sided")$conf
tt<-t.test(two$len, alternative = "two.sided")$conf
list("Half dose interval"=ht,"One dose interval"=ot,"Two dose interval"=tt)
```

```
## $`Half dose interval`
## [1]  8.499046 12.710954
## attr(,"conf.level")
## [1] 0.95
##
## $`One dose interval`
## [1] 17.66851 21.80149
## attr(,"conf.level")
## [1] 0.95
##
## $`Two dose interval`
## [1] 24.33364 27.86636
## attr(,"conf.level")
## [1] 0.95
```

#4. State your conclusions and the assumptions needed for your conclusions.

The p-value of T test for different supplement type is equal to 0.06 and slightly above the 0.05. This tells us that it is not significant or it slightly significant. Confidence interval of OJ is higher and narrower than VC with confidence level of 80%. If we use confidence level of 95% as we do it usual then their intervals intersept, which says there is no 95% confidence in the observations.

However, in comparing the length for different dose amounts, it is stright forward. 0.5 dose length values are less than 1 dose, and 1 dose values are less than 2 dose values with confidence level of 95%.