Part 2

Nirmal Ghimire

10/3/2020

Part 2-Basic Inferential Statistics

This second portion of the project analyzes the ToothGrowth data in the R dataset package and performs the following activities:

1. Basic Exploratory Data Analyses:
2. Basic Summary of the Data:
3. Use Confidence Intervals (CI) and/or hypothesis tests to compare tooth growth by supp and dose:
4. State conclusion and the assumptions needed for conclusion

Invoking Required Libraries

```
library(ggplot2)
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
```

Loading Required Datafile

```
data(ToothGrowth)
```

1. Basic Exploratory Data Analyses

a. Quick Checking of Given Dataset

```
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
```

b. Checking the structure of the dataset

```
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 ...
```

c. Summary of the Dataset

```
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
```

2. Changing dose as a factor

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
```

i. Checking if that worked

```
head(ToothGrowth$dose)
```

```
## [1] 0.5 0.5 0.5 0.5 0.5 0.5
## Levels: 0.5 1 2
```

ii. Verifying the mean of the len variable by supplment types

```
SupMean=split(ToothGrowth$len,ToothGrowth$supp)
sapply(SupMean,mean)
```
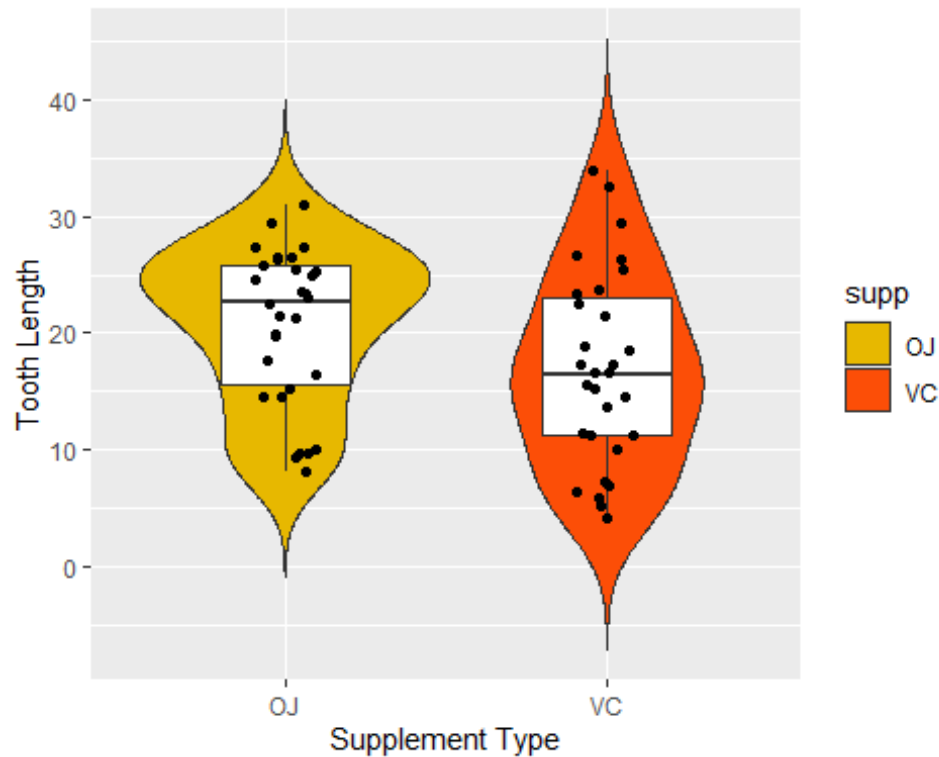
```
##       OJ       VC
## 20.66333 16.96333
```

iv. Creating box plot of OJ and VC

```
e<-ggplot(aes(x=supp,y=len),data=ToothGrowth)
```

Experimenting the violin plot and Including Boxplot within

```
e+geom_violin(aes(fill=supp),trim=FALSE)+
  geom_boxplot(width=0.4)+scale_fill_manual(values=c("#E7B800", "#FC4E07"))+
geom_jitter(width = 0.1)+xlab("Supplement Type")+
  ylab("Tooth Length")
```

Checking the impact of Vitamin C on tooth length

```
Mean_dose=split(ToothGrowth$len,ToothGrowth$dose)
sapply(Mean_dose,mean)

##    0.5     1     2
## 10.605 19.735 26.100
```

Plotting the findings

```
e+ggtitle("Box-Plot Showing Dose of Supplement Type and Tooth Length")+
  geom_boxplot(aes(fill=dose))+scale_fill_manual(values=c("#E7B800",
              "#FC4E07","#00AFBB"))+xlab("Dose")+ylab("Tooth Length")
```

Box-Plot Showing Dose of Supplement Type and Tootl
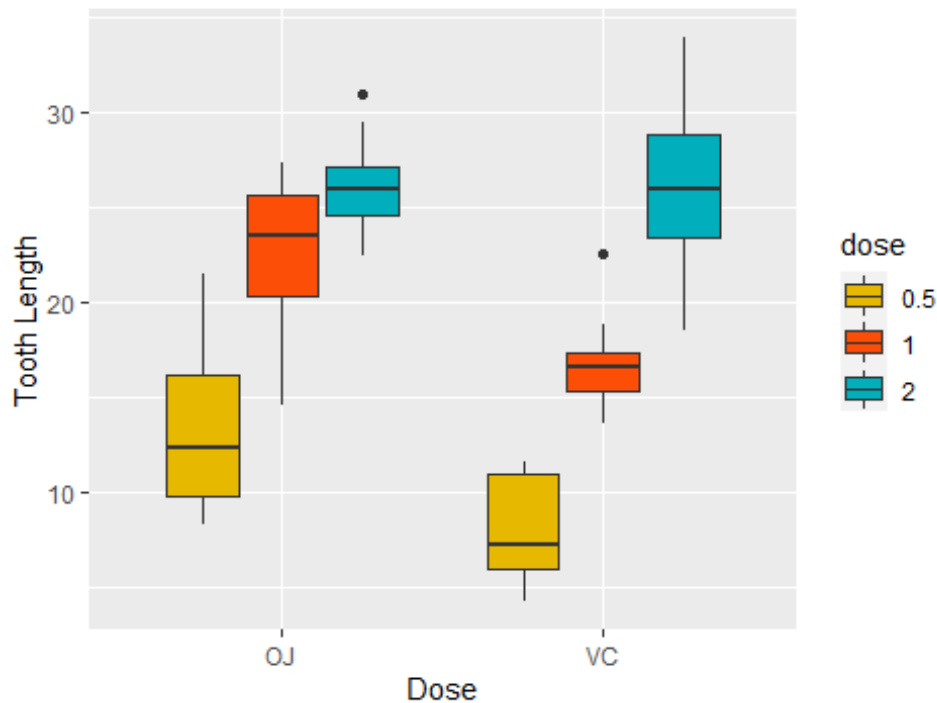
Tooth Length and Delivery Method

```
ToothGrowth %>%
  group_by(supp,dose) %>%
  summarise(Q25th_len=quantile(len,0.25),
            Q50th_len=quantile(len,0.5),
            Q75th_len=quantile(len,0.75),
            average_lenth=mean(len),
            SD_len=sd(len))->newtable
```

```
## `summarise()` regrouping output by 'supp' (override with `.groups`
argument)
```

```
newtable
```

```
## # A tibble: 6 x 7
## # Groups:   supp [2]
##    supp  dose  Q25th_len Q50th_len Q75th_len average_lenth SD_len
##    <fct> <fct>     <dbl>     <dbl>     <dbl>         <dbl>  <dbl>
## 1 OJ    0.5        9.7      12.2      16.2          13.2    4.46
## 2 OJ    1         20.3      23.5      25.6          22.7    3.91
## 3 OJ    2         24.6      26.0      27.1          26.1    2.66
## 4 VC    0.5        5.95      7.15     10.9           7.98   2.75
## 5 VC    1         15.3      16.5      17.3          16.8    2.52
## 6 VC    2         23.4      26.0      28.8          26.1    4.80
```

Calculating ttest to study the relationship between tooth length, supplement type, and dose

```r
test=list()
dose=c(0.5,1,2)
for(m in dose){
  Moj=ToothGrowth$len[ToothGrowth$dose==m & ToothGrowth$supp=="OJ"]
  Mvc=ToothGrowth$len[ToothGrowth$dose==m & ToothGrowth$supp=="VC"]
  t<-t.test(Moj,Mvc)
  id<-paste0("OJ","-","VC",",",m)
  test<-rbind(test,list(id=id,p.value=t$p.value,CI.LOW=t$conf.int[1],
                        CI.HIGH=t$conf.int[2]))
}
test
```

```
##       id             p.value      CI.LOW    CI.HIGH
## [1,] "OJ-VC,0.5" 0.006358607 1.719057 8.780943
## [2,] "OJ-VC,1"   0.001038376 2.802148 9.057852
## [3,] "OJ-VC,2"   0.9638516    -3.79807 3.63807
```