Part 2-Basic Inferential Data Analysis

Harikrishna Vemula

10/13/2020

Load the ToothGrowth data and perform exploratory data analyses

```
library(datasets)
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 ...
```

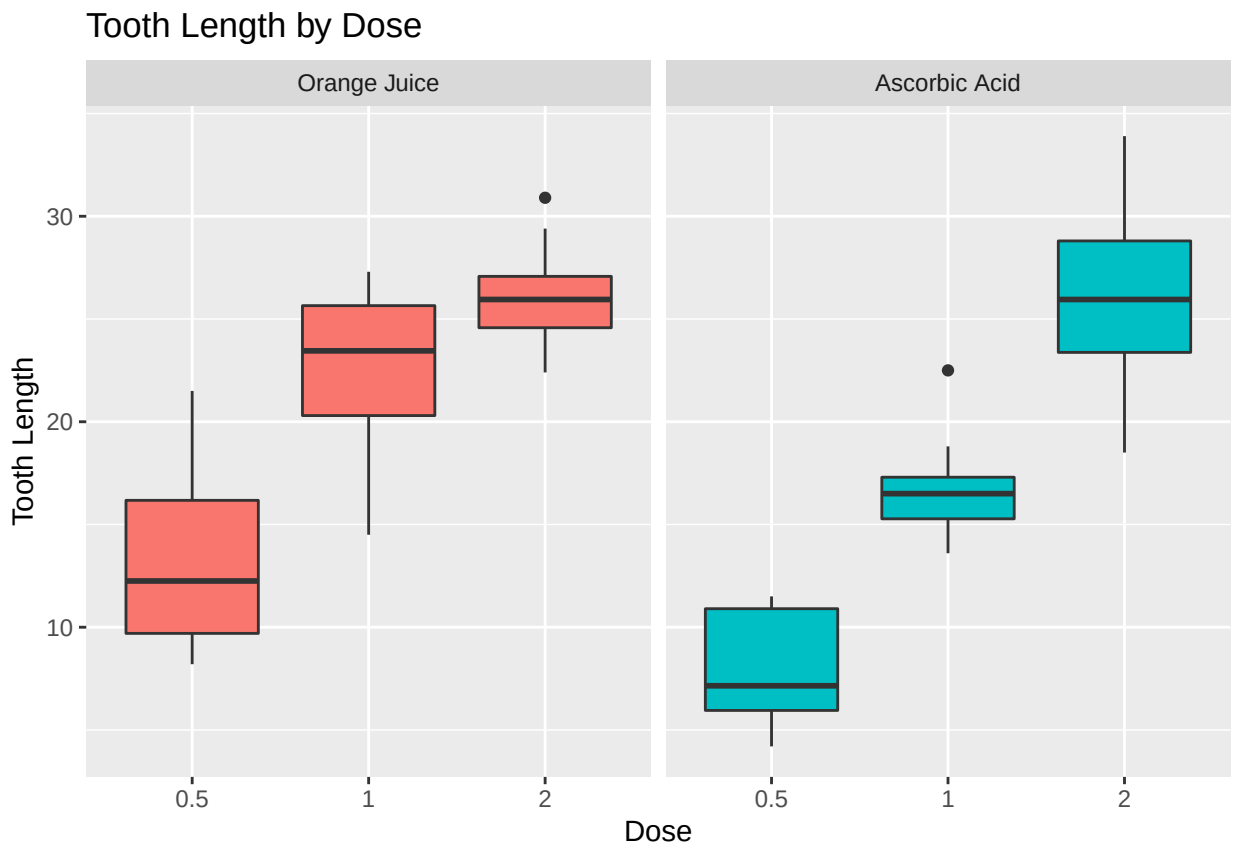
```
head(ToothGrowth)
```

```
##    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
```

```
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
```

```
library(ggplot2)
t = ToothGrowth
levels(t$supp) <- c("Orange Juice", "Ascorbic Acid")
ggplot(t, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +
  geom_boxplot(aes(fill = supp), show.legend = FALSE) +
  labs(title="Tooth Length by Dose",
    x="Dose",
    y="Tooth Length")
```



Basic summary of the data 1. While Dose increases, Tooth Growth is increasing, So Dose is proportional to Tooth Growth 2. Orange Juice is more effective with comapre to Ascorbic Acid, while dose from 0.5 to 2, and when dose is at 2, it slightly similar with Ascorbic Acid.

Use confidence intervals & hypothesis tests to compare tooth growth by supplement and dose

```
test1<-t.test(len ~ supp, data = t)
test1$conf.int
```

Hypo Test:1

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
test1$p.value
```

```
## [1] 0.06063451
```

```
test2<-t.test(len ~ supp, data = subset(t, dose == 0.5))
test2$conf.int
```

Hypo Test:2

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
test2$p.value
```

```
## [1] 0.006358607
```

```
test3<-t.test(len ~ supp, data = subset(t, dose == 1))
test3$conf.int
```

Hypo Test:3

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
test3$p.value
```

```
## [1] 0.001038376
```

```
test4<-t.test(len ~ supp, data = subset(t, dose == 2))
test4$conf.int
```

Hypo Test:4

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
test4$p.value
```

```
## [1] 0.9638516
```

Conclusions & Assumptions

Conclusions

1. While Dose increases, Tooth Growth is increasing, So Dose is proportional to Tooth Growth
2. Orange Juice is more effective with comapre to Ascorbic Acid, while dose from 0.5 to 2, and when dose is at 2, it slightly similar with Ascorbic Acid.

Assumptions 1.By assuming Hypo tests the underlying data is independent and normally distributed. 2.Observed that the data represents random id samples.