Statistical Inference Part 2

Nur Seto Dimas

15 September 2019

Overview

The second part of the project will performs analysis toward Tooth Growth form R data set package. From the course instructions, these following steps will be used to analyze the data set.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Analysis

Exploratory Data Analysis

```
library(ggplot2)
data("ToothGrowth")     # load the data set

# Summary of the data set
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
```

```
# Views data set structure
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 ...
```

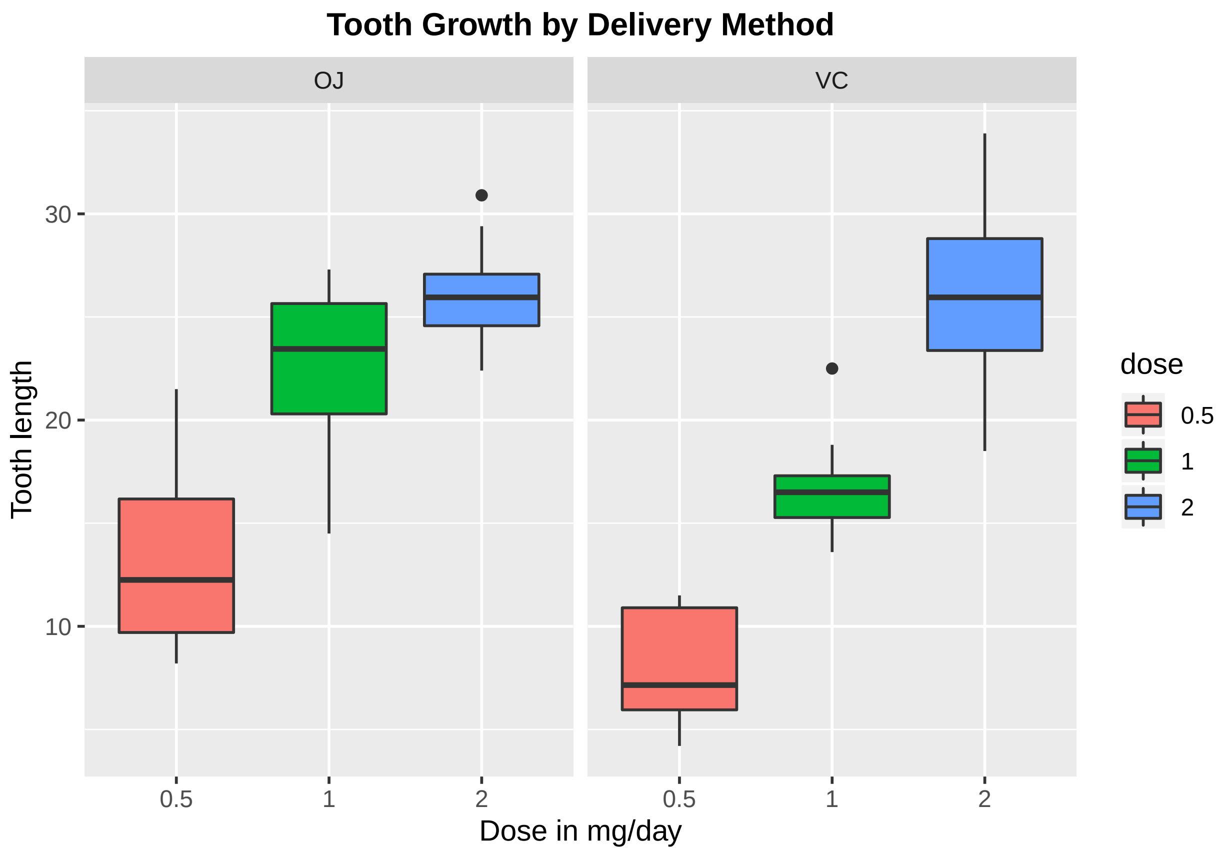
There are three variables in the data set:

1. len - growth in tooth length
2. supp - delivery method indicated by type of supplement provided (VC - ascorbic acid, a form of vitamin C; OJ - orange juice)
3. dose - milligrams / day of vitamin C provided

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

ggplot(data = ToothGrowth, aes(x = dose, y = len)) +
        geom_boxplot(aes(fill = dose)) +
        facet_grid( . ~ supp) +
```

```r
    ggtitle(label = "Tooth Growth by Delivery Method") +
    theme(plot.title = element_text(size = 12, hjust = 0.5, face = "bold")) +
    xlab("Dose in mg/day") +
    ylab("Tooth length")
```



The box plot shows positive relation between length of tooth growth with dose amounts. Higher dose will result higher growth in tooth's length. The delivery methods shown similar pattern with the increase amount of dose. Higher dose in each delivery method provide better result in tooth growth.

Hypothesis Testing

Delivery Method Impact

```r
ttest_all <- t.test(data = ToothGrowth, len~supp)$p.value

dose_0.5 <- subset(ToothGrowth, dose == 0.5)
ttest_0.5mg <- t.test(data = dose_0.5, len~supp)$p.value

dose_1 <- subset(ToothGrowth, dose == 1)
ttest_1mg <- t.test(data = dose_1, len~supp)$p.value

dose_2 <- subset(ToothGrowth, dose == 2)
ttest_2mg <- t.test(data = dose_2, len~supp)$p.value

rbind(ttest_all, ttest_0.5mg, ttest_1mg, ttest_2mg)
```

```
##                    [,1]
## ttest_all   0.060634508
## ttest_0.5mg 0.006358607
## ttest_1mg   0.001038376
## ttest_2mg   0.963851589
```

The null hypothesis for this test is there is no significant effect from delivery method (supp) on tooth growth.

The p-value from the t test resulted higher than 0.5 (**0.060634508**). So the null hypothesis is accepted. However, if further observation performed, the p-value for each dose hold different values.

1. For 0.5 mg/day dosage, the p-value is **0.006358607**, lower than 0.5. Thus null hypothesis can be rejected, there is significant effect in tooth growth from delivery method with 0.5 mg/day dose.
2. For 1 mg/day dosage, with p-value **0.001038376**, same conclusion can be made. There is significant effect on tooth growth from delivery method.
3. For 2 mg/day dosage, the p-value is 0.963851589, higher than 0.5. The null hypothesis can't be rejected.

Dose Amounts Impact

```r
# for dose 0.5 and 1.0 mg/day
dose_1 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1.0))
ttest_dose1 <- t.test(data = dose_1, len~dose)$p.value

# for dose 0.5 and 2.0 mg/day
dose_2 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
ttest_dose2 <- t.test(data = dose_2, len~dose)$p.value

# for dose 1.0 and 2.0 mg/day
dose_3 <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))
ttest_dose3 <- t.test(data = dose_3, len~dose)$p.value

rbind(ttest_dose1, ttest_dose2, ttest_dose3)
```

```
##                     [,1]
## ttest_dose1 1.268301e-07
## ttest_dose2 4.397525e-14
## ttest_dose3 1.906430e-05
```

All p-values are smaller than 0.5 so the null hypothesis is rejected There is significant effect from dose amounts to tooth length's growth.

Conclusion

Assumption

1. The dependent variable (len) is normally distributed
2. The variance between two group of mean (len) are equal.
3. Every observation conducted are independent to each other.

Conclusion

1. The initial test of supplement resulted that it has no impact on tooth growth. However with further analysis, the dose 0.5 and 1.0 mg/day for each supplement has significant effect on tooth growth while 2.0 mg/day is the opposite. These results also supported with boxplot from EDA section, showing increase in tooth growth with increasing dose on each type of supplement.
2. From exploratory data analysis and hypothesis test showed that dosage level has significant effect on tooth growth. The higher the dosage level, the higher the tooth length.