

Statistical Inference Project Pt 2

Hsin Chih Chen

10/6/2021

Introduction

The ToothGrowth data in R datasets will be analyzed accordingly to find the data summary and use hypothesis tests to compare the tooth growth factor by either the supplement type and dose type.

Preliminary Setup & Store Data

```
# load graphing tool interface package
library(ggplot2)

# initialize data category
Tooth <- ToothGrowth
```

Provide Initial Summaries for the data structure.

```
# General summaries for the data frame
summary(Tooth)
```

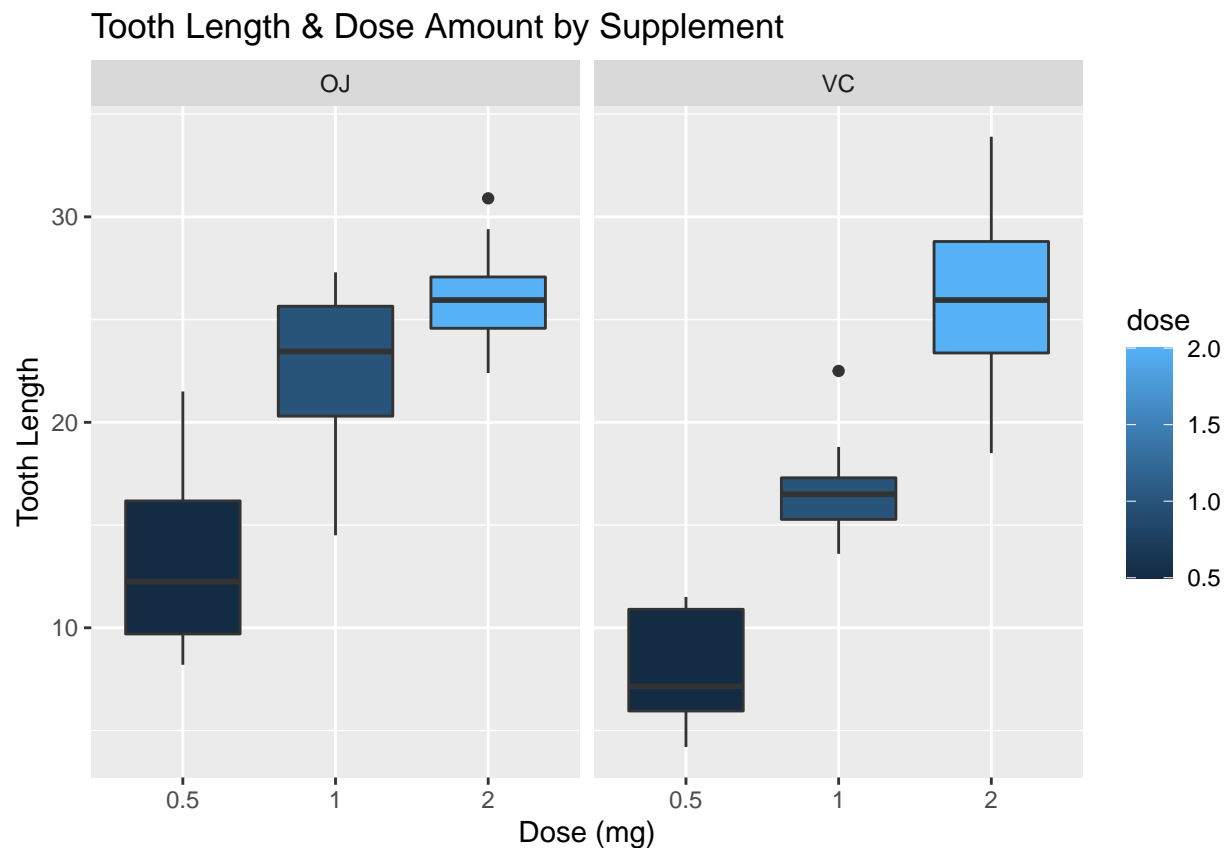
```
##      len      supp      dose
##  Min.   : 4.20   OJ:30   Min.    :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.    :2.000
```

```
# Examine the characteristics for each section
str(Tooth)
```

```
## '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 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

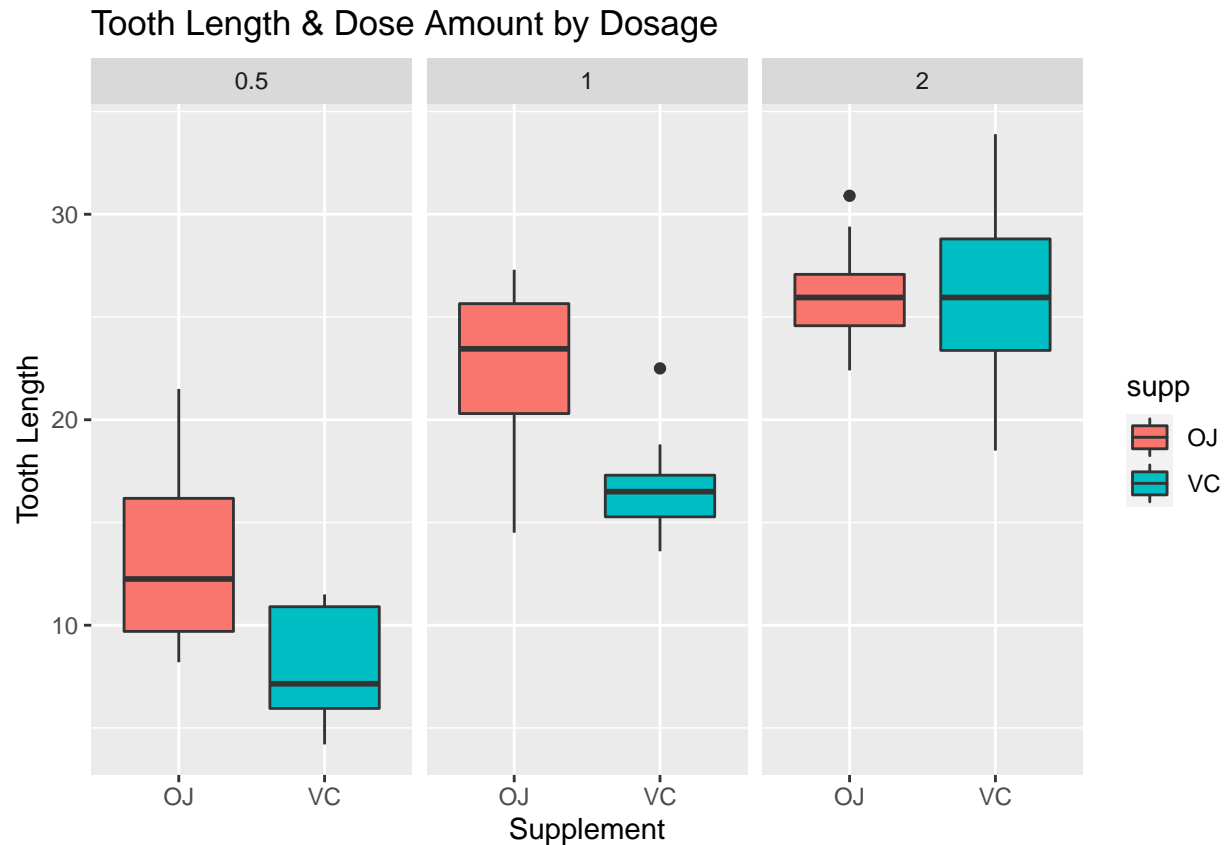
Plot Tooth grow length and dose by splitting the supplement

```
# Use boxplot by having the classified grid for supplement
ggplot(aes(x = as.factor(dose), y = len), data = Tooth) +      geom_boxplot(aes(fill = dose)) +
  xlab("Dose (mg)") +
  ylab("Tooth Length")+
  facet_grid(~ supp) +
  ggtitle("Tooth Length & Dose Amount by Supplement")
```



Plot Tooth grow length and supplement by splitting the dosage amount

```
# Use boxplot by having the classified grid for dosage
ggplot(aes(x = supp, y = len), data = Tooth) +
  geom_boxplot(aes(fill = supp)) +
  xlab("Supplement") +
  ylab("Tooth Length") +
  facet_grid(~ dose) +
  ggtitle("Tooth Length & Dose Amount by Dosage")
```



Have Initial Hypothesis Test for the supplement classification

In this Hypothesis testing, the following hypothesis test is listed below:

H_0 = Both supplement will generate the same impact towards tooth length growth. H_1 = Both supplement will generate different impact towards tooth length growth.

```
# Use 2 sample t-test to compare
t.test(len ~ supp, data = Tooth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##      20.66333      16.96333
```

Have Initial Hypothesis Test for the dosage classification

In this Hypothesis testing, the following hypothesis test is listed below:

H0 = Both dosage will generate the same impact towards tooth length growth. H1 = Both dosage will generate different impact towards tooth length growth.

```
# Use 2 sample t-test to compare where the dosage is 0.5 and 1
Tooth1 <- subset(Tooth, Tooth$dose %in% c(0.5,1))
t.test(len ~ dose, data = Tooth1)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
## -11.983781 -6.276219
## sample estimates:
## mean in group 0.5 mean in group 1
## 10.605 19.735
```

```
# Use 2 sample t-test to compare where the dosage is 1 and 2
Tooth2 <- subset(Tooth, Tooth$dose %in% c(1,2))
t.test(len ~ dose, data = Tooth2)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
## -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
## 19.735 26.100
```

```
# Use 2 sample t-test to compare where the dosage is 0.5 and 2
Tooth3 <- subset(Tooth, Tooth$dose %in% c(0.5,2))
t.test(len ~ dose, data = Tooth3)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

Statistical Conclusion

Based on the t-test results and p-value description, the following conclusions are made.

1. Supplement type will NOT contribute different impact towards the tooth length based on the given p-value.
2. Dosage type will contribute different impact towards the tooth length based on the given p-value across all 3 factors of dosage. (H_0 was rejected)