

Statistical Inference Course Project Part 2

Keng Yew Hoe

10/13/2020

Overview

Analyzing the ToothGrowth data in the R datasets package while providing a basic summary of the data and using confidence intervals with hypothesis tests to compare tooth growth by supp and dose.

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

```
# Load the ToothGrowth data
```

```
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
```

```
# Perform Exploratory Data Analysis
```

```
library(ggplot2)
```

```
#Replacing OJ and VC with Orange Juice and Vitamin C
```

```

ToothGrowth$supp <- as.character(ToothGrowth$supp)
ToothGrowth$supp[ToothGrowth$supp == "OJ"] <- "Orange Juice"
ToothGrowth$supp[ToothGrowth$supp == "VC"] <- "Vitamin C"

```

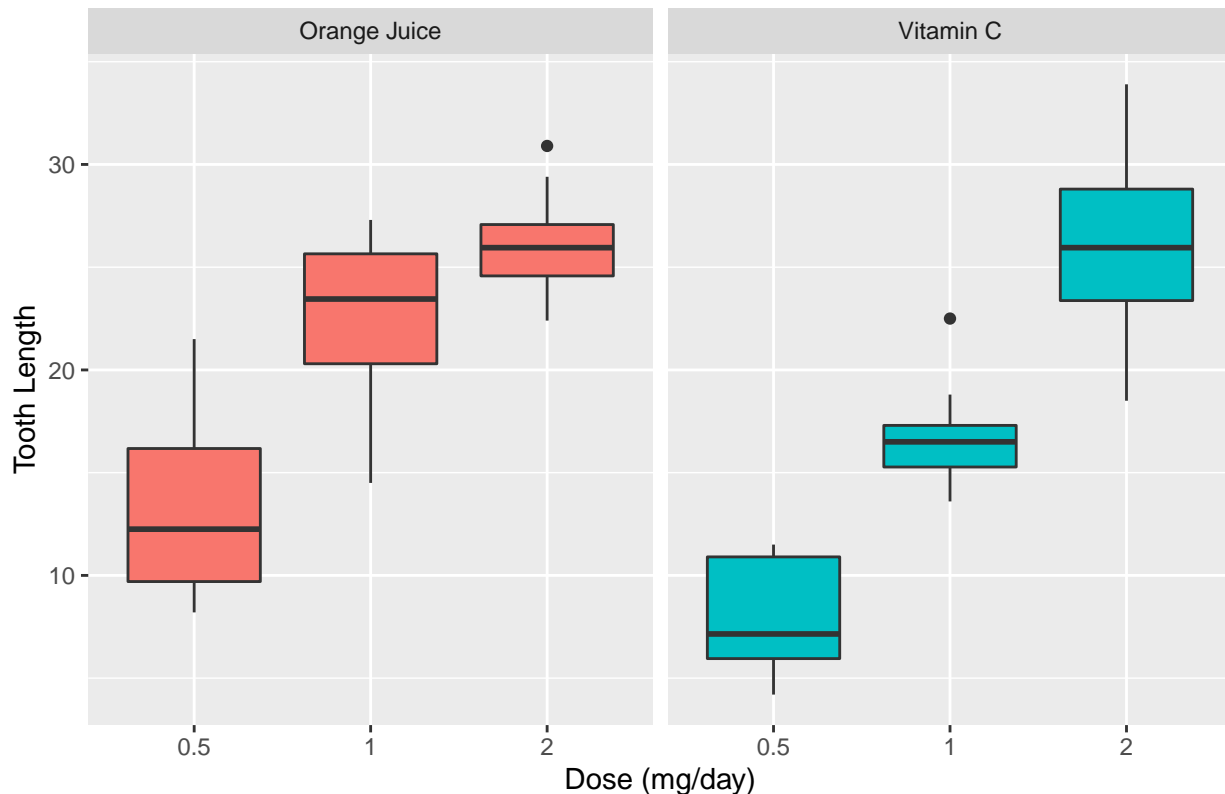
```

#Plotting
ggplot(ToothGrowth, aes(x=factor(dose), y=len)) +
  facet_grid(~supp) +
  geom_boxplot(aes(fill = supp), show_guide = FALSE) +
  labs(title="Tooth length along dosage for each type of supplement",
       x="Dose (mg/day)",
       y="Tooth Length")

```

Warning: 'show_guide' has been deprecated. Please use 'show.legend' instead.

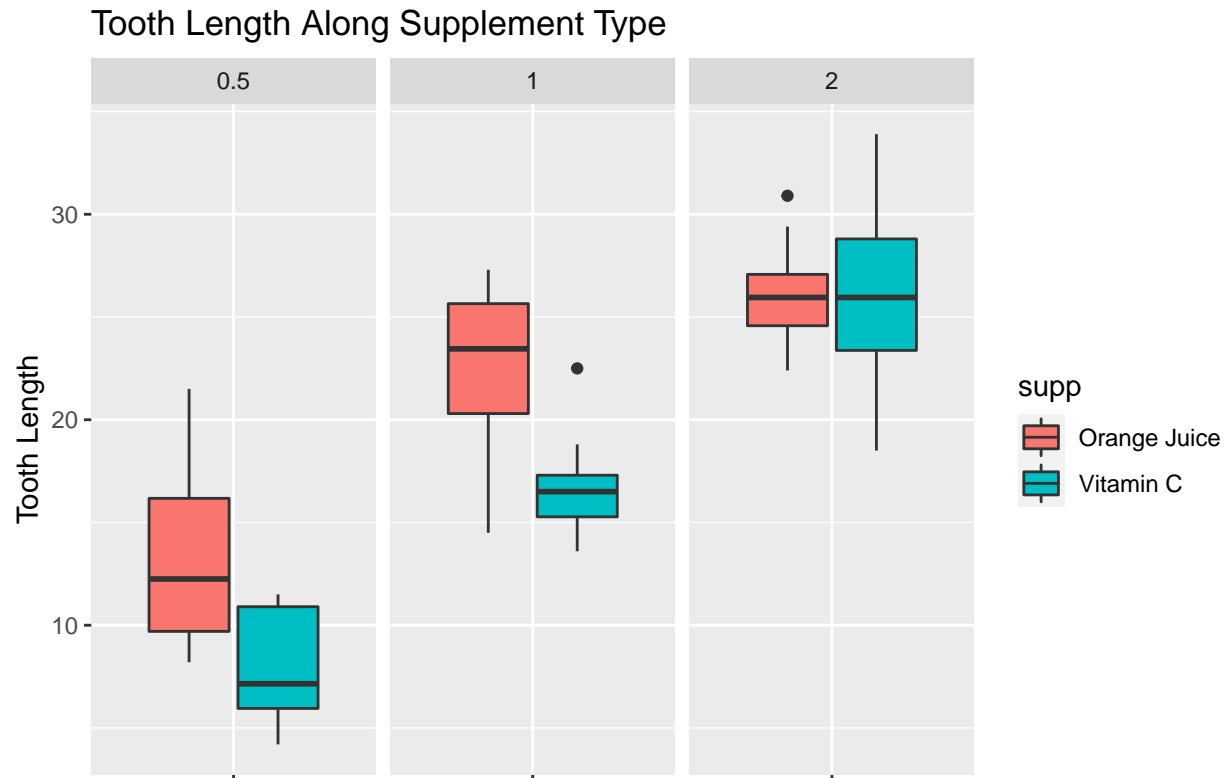
Tooth length along dosage for each type of supplement



```

g <- ggplot(aes(x = " ", y = len), data = ToothGrowth)
g + geom_boxplot(aes(fill=supp))+ xlab(" ") +
  ylab("Tooth Length") + facet_grid(~ dose) + ggtitle("Tooth Length Along Supplement Type")

```



2. Basic summary of the data

Based on the graph, tooth growth increases as dosage increases. It is also shown that orange juice has a higher median tooth length when dosage is 0.5 and 1 mg/day. Both types of supplements have the same median when dosage is 2 mg/day.

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

Hypothesis 1

Orange juice & Vitamin C deliver the same tooth growth across the data set.

```
h1<-t.test(len ~ supp, ToothGrowth)
h1$conf.int
```

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

```
h1$p.value
```

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

The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05. The null hypothesis cannot be rejected.

Hypothesis 2

For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth.

```
h2<-t.test(len ~ supp, subset(ToothGrowth, dose == 0.5))
h2$conf.int
```

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

```
h2$p.value
```

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

The confidence interval does not include 0 and the p-value is below the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 0.5 mg/day dosage of orange juice delivers more tooth growth than Vitamin C is accepted.

Hypothesis 3

For the dosage of 1 mg/day, the two supplements deliver the same tooth growth

```
h3<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
h3$conf.int
```

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

```
h3$p.value
```

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

The confidence interval does not include 0 and the p-value is smaller than the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 1 mg/day dosage of orange juice delivers more tooth growth than Vitamin C is accepted.

Hypothesis 4

For the dosage of 2 mg/day, the two supplements deliver the same tooth growth

```
h4<-t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
h4$conf.int
```

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

```
h4$p.value
```

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

The confidence interval does include 0 and the p-value is larger than the 0.05 threshold. The null hypothesis cannot be rejected.

State your conclusions and the assumptions needed for your conclusions

Orange juice delivers more tooth growth than Vitamin C for dosages 0.5 & 1.0 mg/day. Orange juice and Vitamin C deliver the same amount of tooth growth for dose amount 2.0 mg/day. For the entire data set we cannot conclude orange juice is more effective than vitamin C however it can be concluded that dosage will cause an increase in tooth length

Assumptions

- Normal distribution of the tooth lengths
- No other unmeasured factors are affecting tooth length