Statistical Inference Course Project Part 2 : Analyze the ToothGrowth Dataset

Jonathan Lok-Chuen Lo

October 25, 2015

Overview

In this project, I will analyze the ToothGrowth data in the R datasets package. Basic exploratory data analyses will be performed. Confidence intervals and hypothesis tests are also used to compare tooth growth by supp and dose. Finally an conclusions and the corresponding assumptions will be given.

Load the Data

Load the ToothGrowth data in the dataset package, and all the necessary libraries. Convert dose into factor for plotting purposes later.

```
library(datasets)
library(lattice)

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

Basic Summary of the Data

Display the structure and summary of the dataset.

```
str(ToothGrowth)
```

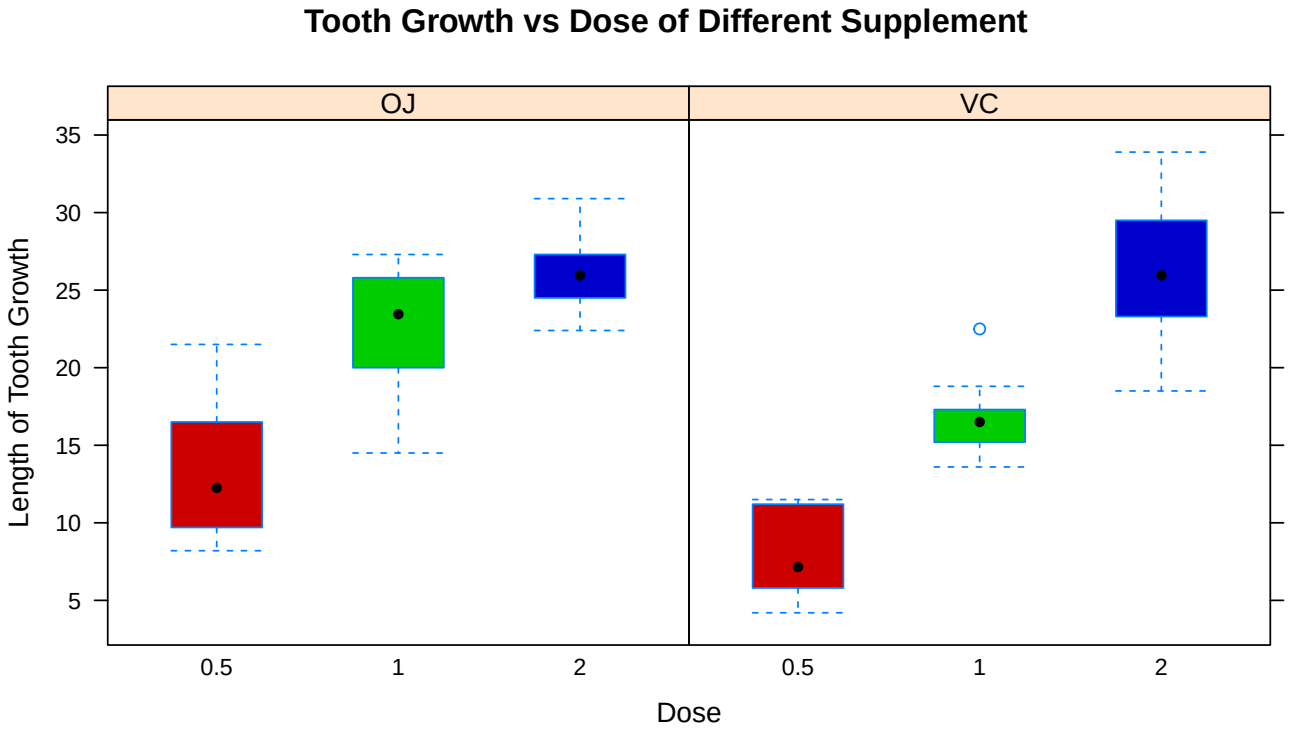
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(ToothGrowth)
```

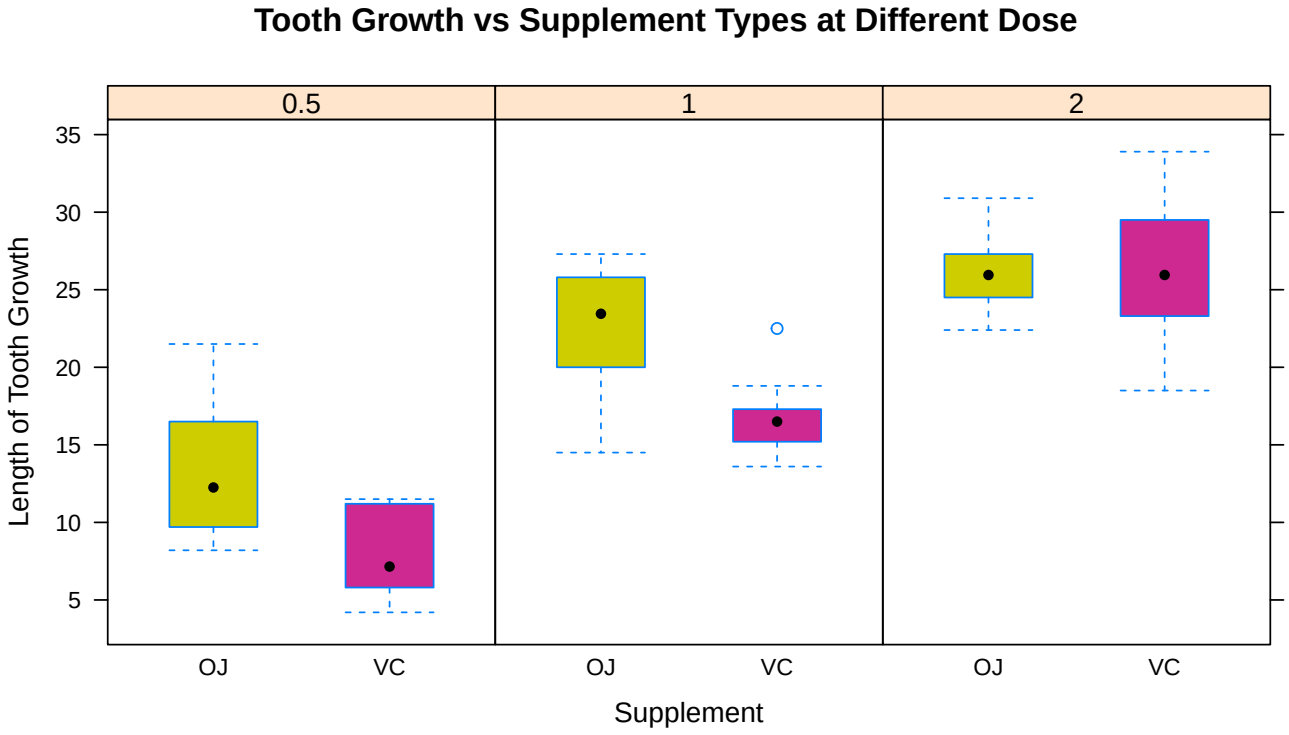
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Exploratory Data Analysis

Plot the tooth growth against dose for different supplement type. Increase dosage seems to stimulate faster tooth growth for both supplement types.



Plot the tooth growth against supplement type at different dosage. The effect is not obvious.



Compare Tooth Growth by Supplement and Dose

Perform t-test on tooth growth between different supplement types.

```
t <- t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"],ToothGrowth$len[ToothGrowth$supp == "VC"],
            paired = FALSE, var.equal = FALSE)
list(PValue = t$p.value, ConfidenceInterval = as.numeric(t$conf.int))
```

```
## $PValue
## [1] 0.06063451
##
## $ConfidenceInterval
## [1] -0.1710156  7.5710156
```

P-value is 6.06%. This indicates we can accept the null hypothesis of difference in tooth growth being zero using different supplement with 95% confidence. This is also shown that zero is included within the 95% confidence interval.

Perform t-test on tooth growth when dosage increase from 0.5mg to 2mg.

```
t <- t.test(ToothGrowth$len[ToothGrowth$dose == "2"], ToothGrowth$len[ToothGrowth$dose == "0.5"],
       paired = FALSE, var.equal = FALSE)
list(PValue = t$p.value, ConfidenceInterval = as.numeric(t$conf.int))
```

```
## $PValue
## [1] 4.397525e-14
##
## $ConfidenceInterval
## [1] 12.83383 18.15617
```

P-value is 0.00%. This indicates we can reject the null hypothesis, i.e. accept the alternative of difference in tooth growth not equal to zero as dosage increases. The 95% confidence interval does not include zero also means that tooth growth is increased by increasing the dosage from 0.5mg to 2mg.

Conclusions

Using the t-test analysis, we have concluded the following:

1. Supplement type has no statistically significant effect on tooth growth.
2. Tooth growth can be increasing by increasing the dosage level.

The analysis is based on the following assumptions:

- Dosage and supplement types are random assigned to the guinea pigs.
- Sample population are representative of the entire population of guinea pigs.
- For the t-tests, data area assumed not paired and variances are different among different groups.