Statistical Inference Course Project: 2nd Part

Dineshkumar Murugan

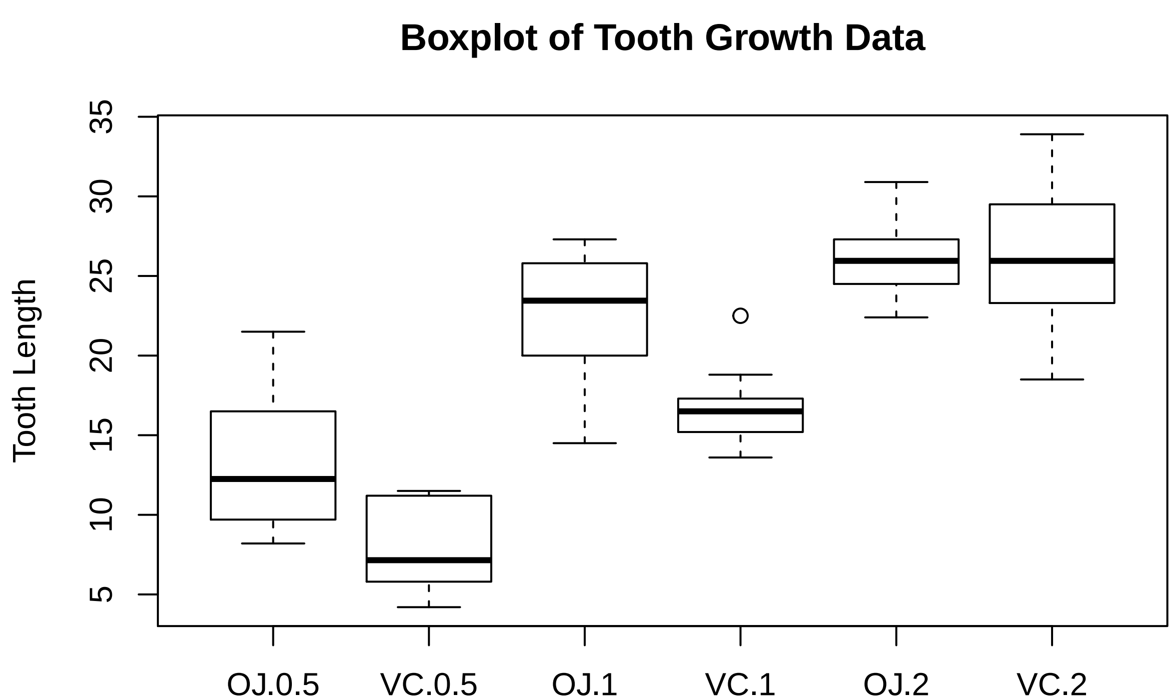
Saturday, May 23, 2015

Load the Dataset

```
library(datasets)
data(ToothGrowth)
```

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

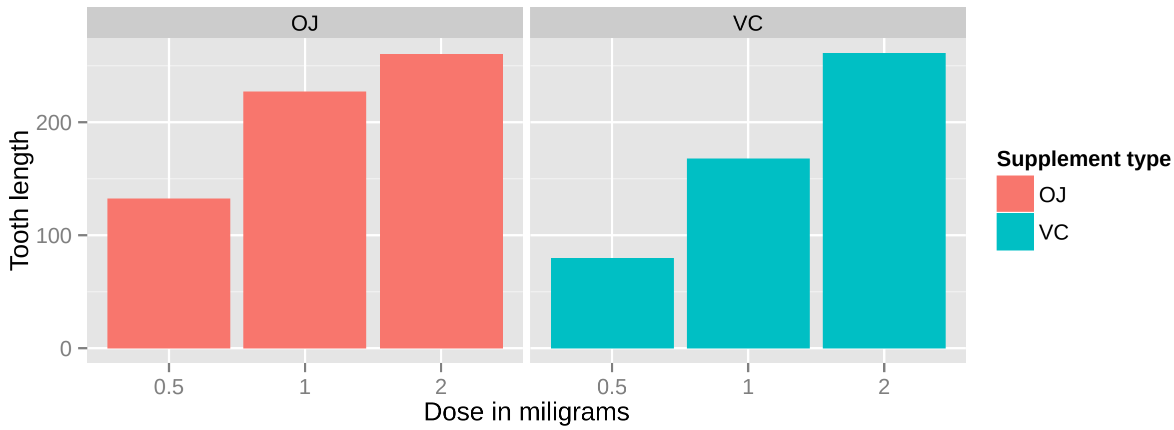
```
boxplot(len ~ supp * dose, data=ToothGrowth, ylab="Tooth Length", main="Boxplot of Tooth Growth Data")
```



As we can see, on average the length of the tooth increases as the dose also increases

```
library(ggplot2)
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity",) +
    facet_grid(. ~ supp) +
    xlab("Dose in miligrams") +
```

```
    ylab("Tooth length") +
    guides(fill=guide_legend(title="Supplement type"))
```



As can be seen above, there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

2. Provide a basic summary of the data.

```
library(plyr)
summaryDT <- ddply(ToothGrowth,.(dose, supp), summarize, mean = mean(len), sd = sd(len))
as.factor(summaryDT$dose)
```

```
## [1] 0.5 0.5 1   1   2   2
## Levels: 0.5 1 2
```

```
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
```

```
nrow(ToothGrowth)
```

```
## [1] 60
```

```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##      
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

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)

The degree of freedom is more than 30 we can go for Z distribution for calculating confidence interval. But it will not hurt to use T distribution as both will give a near answer.

The confidence interval is for the difference between the length of toothlength under the effect of OJ and VC for each dose.

```
ToothGrowth <- transform(ToothGrowth, dose=as.factor(dose))
s <- split(ToothGrowth, list(ToothGrowth$supp, ToothGrowth$dose))
t1 <- t.test(s[[1]][[1]], s[[2]][[1]], paired = TRUE, alternative = "greater",conf.level = 0.95)
t2 <- t.test(s[[3]][[1]], s[[4]][[1]], paired = TRUE, alternative = "greater",conf.level = 0.95)
t3 <- t.test(s[[5]][[1]], s[[6]][[1]], paired = TRUE, alternative = "greater",conf.level = 0.95)
```

1. The confidence level for dose 0.5 is 2.0196, 8, and the p-value is 0.0077
2. The confidence level for dose 0.5 is 2.7064, 8, and the p-value is 0.0041
3. The confidence level for dose 0.5 is -3.5231, 8, and the p-value is 0.5165

4. State your conclusions and the assumptions needed for your conclusions.

Assumptions: H0: both supp have the same impact on tooth length. Ha: OJ can promote the growth of toothlength than VC. significance level: 0.05

Conclusions: OJ have more impact on the growth of toothlength under dose equals to 0.5 or 1.But can't get this conclusion when dose equals to 2. As the doses increases growth also increases. But they don't seem to be related, we can clearly say based on the given data delivery methods are independent of the dose size.