Exploratory Data Analysis of Tooth Growth Dataset

Ritesh Kumar Malaiya

Saturday, May 23, 2015

Lets look at the data

```
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
```

By looking at the Summary of data we can identify that we would require ***Multi-Variate analysis*** for this data. Also, it may not be correct to analyse Mean, Variance etc without seperating data for VC and OJ.

Summarize data separetely for VC and OJ

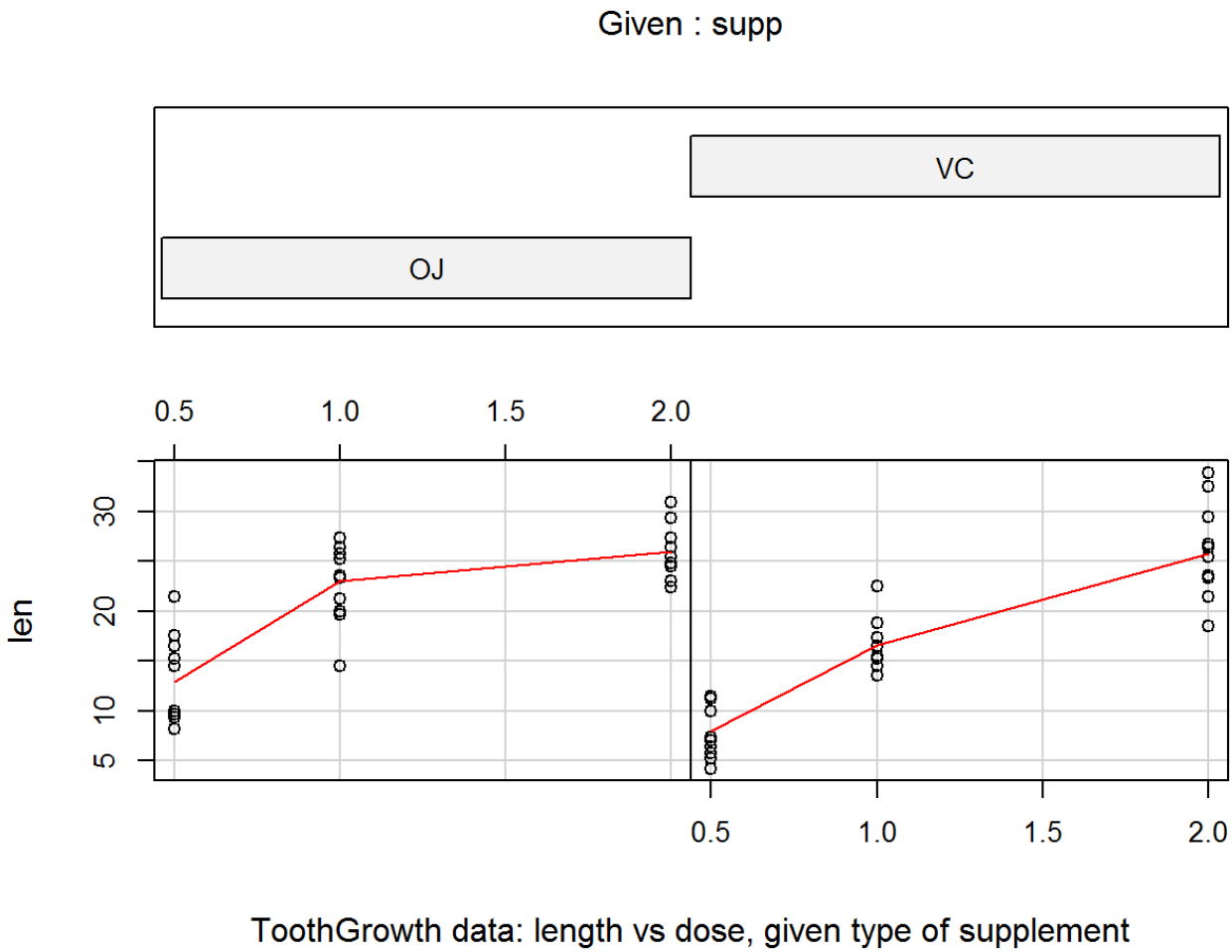
So that Summary would make more sense.

```
library(plyr)
summ <- ddply(ToothGrowth, "supp", summarise, min = min(len), max = max(len), median = me
dian(len), quant.first = quantile(len, probs = c(0)), quant.third = quantile(len, probs =
c(.75)))
rownames(summ) <- summ[,1]
summ[,1] <- NULL
summ
```

```
##    min  max median quant.first quant.third
## OJ 8.2 30.9   22.7         8.2      25.725
## VC 4.2 33.9   16.5         4.2      23.100
```

Let's compare variation in results for dosage in case of VC and OJ

```
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth, xlab = "ToothGrowth d
ata: length vs dose, given type of supplement")
```



This dataset looks like paired (*if we assume that both experiments are done on same guinea pigs*), hence we will use Paired t-test (VC and OJ) to find Confidence intervals/ hypothesis tests.

Paired T-Tests

```
ToothGrowth.vc <- ToothGrowth[ToothGrowth$supp == 'VC','len']
ToothGrowth.oj <- ToothGrowth[ToothGrowth$supp == 'OJ','len']

t.test(ToothGrowth.oj, ToothGrowth.vc, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  ToothGrowth.oj and ToothGrowth.vc
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.408659 5.991341
## sample estimates:
## mean of the differences
##                     3.7
```

Conclusions.

Assumption: *Both experiments are done on same guinea pigs* Based on above analysis, we can conclude that We get: **better results if we increase dose from 0.5 to 1.0 while using Orange Juice as Delivery Method**. However, **if we use Ascorbic Acid and increase dose from 1.0 to 2.0 we get almost same results as Orange Juice**.

Also, ***T-Tests shows that overall growth while using Orange Juice is better than Ascorbic Acid***.