

Tooth Growth Data Investigation

Zhongyi Lin

Aug 17, 2015

Overview

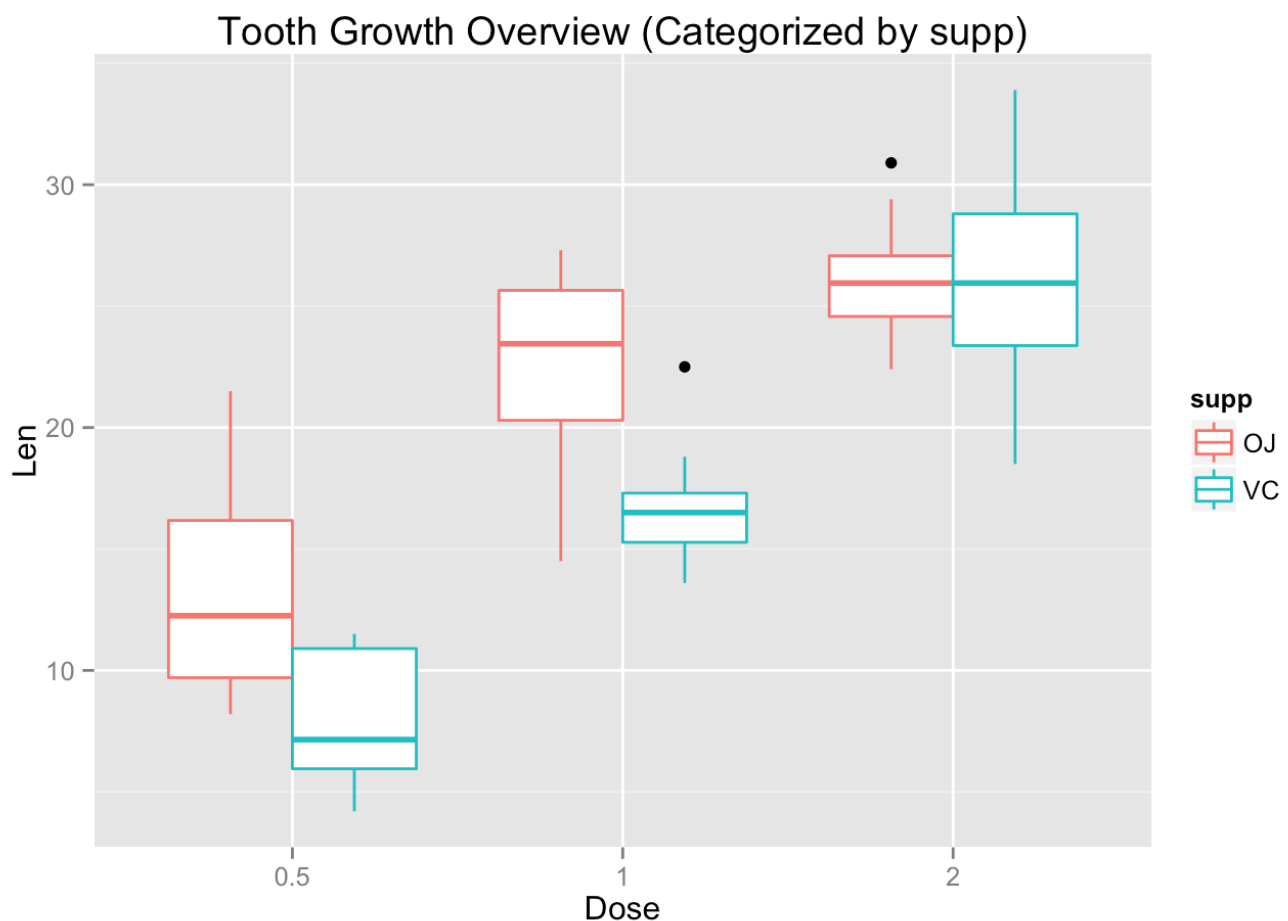
In the second part of the project, we investigate the ToothGrowth data set and compare tooth growth by the supp and dose variables. Figures aided for understanding are plotted using the ggplot2 plotting system.

Exploratory Data Analysis

Firstly the ToothGrowth dataset and ggplot2 package are loaded for use. The code is hidden here.

We can have a direct and intuitive overview on the data by plotting the length variable versus supp and dose variable, as shown in the following two figures.

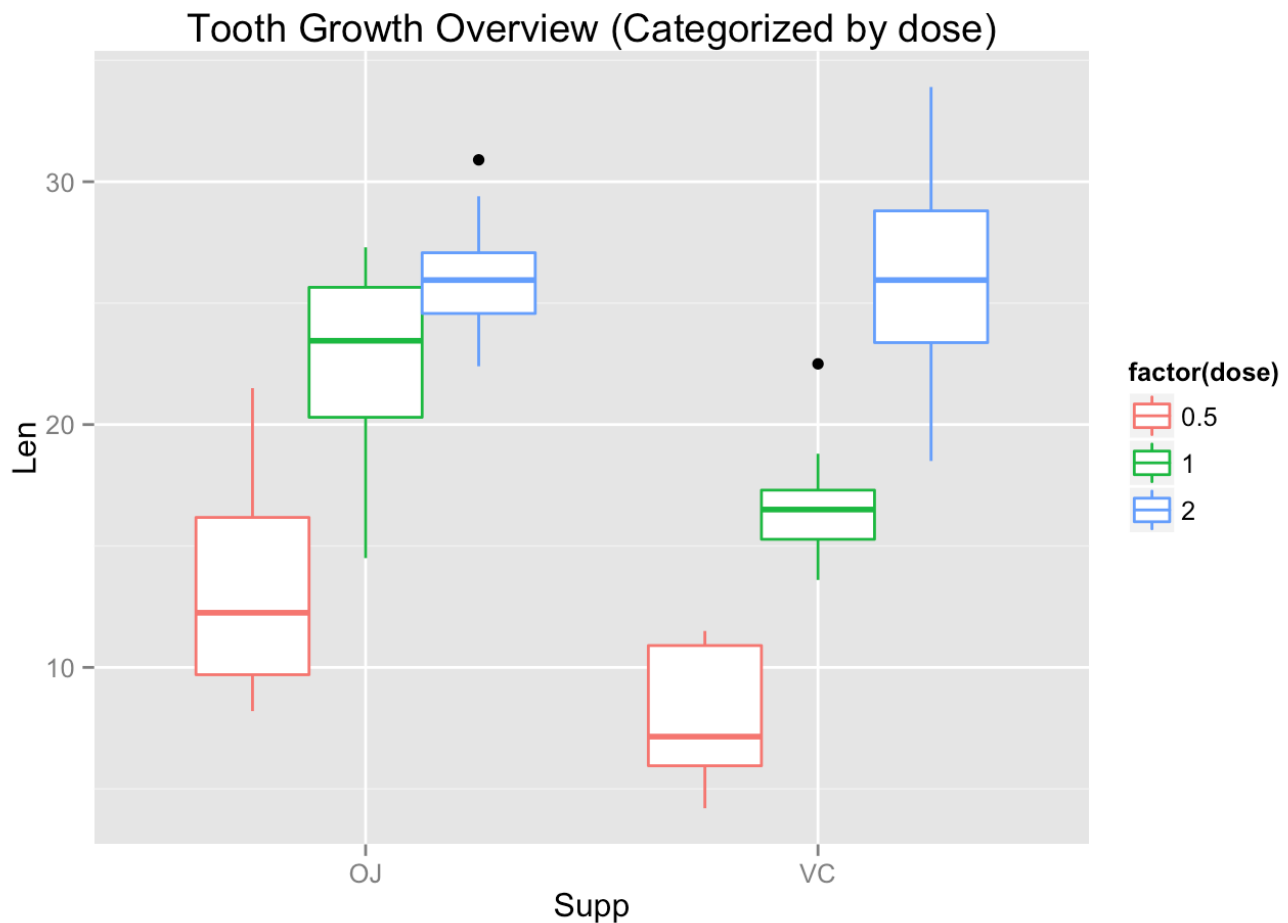
```
tgImg1 <- ggplot(ToothGrowth, aes(x=factor(dose), y=len)) +  
  geom_boxplot(aes(color=supp)) +  
  xlab("Dose") + ylab("Len") + ggtitle("Tooth Growth Overview (Categoriz  
ed by supp)")  
tgImg1
```



```

tgImg2 <- ggplot(ToothGrowth, aes(x=supp, y=len)) +
  geom_boxplot(aes(color=factor(dose))) +
  xlab("Supp") + ylab("Len") + ggtitle("Tooth Growth Overview (Categoriz
ed by dose)")
tgImg2

```



And the summary of the data set is also given here:

```
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

```

Comparison of tooth growth using confidence interval and T test

The data set is splitted by supp variable first. Then we use the two sided 95% confidence interval of the mean difference of OJ and VC as following:

```
OJ <- split(ToothGrowth, ToothGrowth$supp)[[1]]$len
VC <- split(ToothGrowth, ToothGrowth$supp)[[2]]$len
df <- length(OJ) + length(VC) - 2
sp <- sqrt(((length(OJ)-1)*sd(OJ)^2+(length(VC)-1)*sd(VC)^2)/df)
interval <- mean(OJ) - mean(VC) + c(-1,1)*qt(0.975, df)*sp*sqrt(1/length(OJ)+1/length(VC))
interval
```

```
## [1] -0.1670064  7.5670064
```

It can be observed that it is quite confidence for one to say the mean tooth growth of OJ is larger than that of VC. To make sure we calculate the probability of the mean difference to be larger than 0, which is:

```
pt(-(0 - mean(OJ) + mean(VC))/sp/sqrt(1/length(OJ)+1/length(VC)), df)
```

```
## [1] 0.9698033
```

It correspondes to the result we see in the figure, meaning that tooth growth of OJ is larger than that of VC.

It is interesting to particularly investigate the tooth growth of dose 2.0, as we can see the OJ and VC means are quite close. We apply a T test on the data, with the assumption that the null hypothesis is true:

```
temp <- split(ToothGrowth, ToothGrowth$dose)[[3]]
OJ2 <- split(temp, temp$supp)[[1]]
VC2 <- split(temp, temp$supp)[[2]]
t.test(OJ2$len-VC2$len)
```

```
##
## One Sample t-test
##
## data:  OJ2$len - VC2$len
## t = -0.042592, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -4.328976  4.168976
## sample estimates:
## mean of x
## -0.08
```

Here we see a symmetrical 95% confidence interval and a large p value, which means that the null hypothesis of $\mu=0$ is not tend to be rejected. This conclusion correspondes to what we see in the figure.

We also compare the tooth growth by the dose variable. To simplify we just compare dose of 1 and 2 which are closer to each other and giving more possibility. The assumption is the null hypothesis is true.

```
oneZero <- split(ToothGrowth, ToothGrowth$dose)[[2]]$len  
twoZero <- split(ToothGrowth, ToothGrowth$dose)[[3]]$len  
t.test(twoZero-oneZero)
```

```
##  
## One Sample t-test  
##  
## data: twoZero - oneZero  
## t = 4.6046, df = 19, p-value = 0.0001934  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 3.471814 9.258186  
## sample estimates:  
## mean of x  
## 6.365
```

The result obviously shows that tooth growth of dose 2.0 has a mean larger than that of dose 1.0. The p value is quite small thus the null hypothesis could be rejected confirmly.