

PA_ToothGrowth data analysis

wendy

January 24, 2015

Overview: This is basic data analysis for the ToothGrowth. It includes the basic dataset analysis and also the data comparison between supplies and doses.

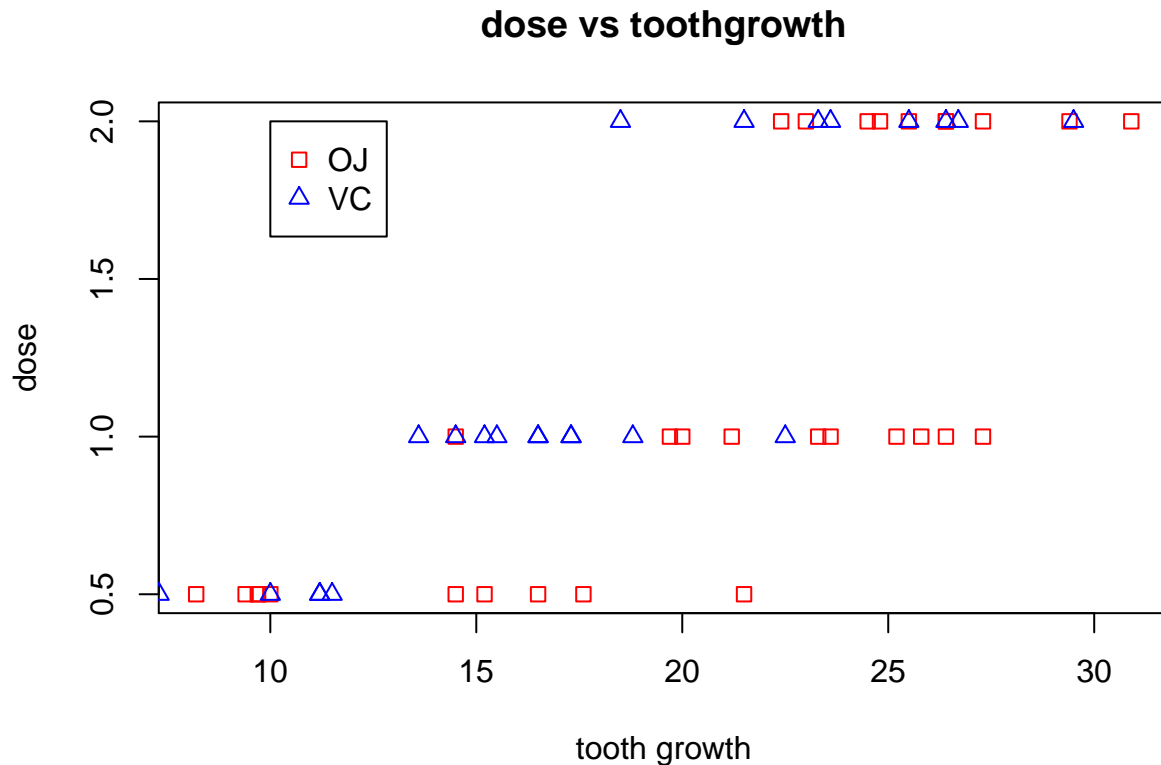
Datasets Basic Analysis: Following code is to restructure the data and show the basic data information. The new dataframe, TG_new, showing the mean tooth growth length by different dose and supplies. I used the IQR to represent the data variation, because IQR could show the data variation more accurate when the data distribute is not symmetric and sample size is not big enough.

```
library(datasets); data(ToothGrowth)
supp_l<-levels(ToothGrowth$supp); dose_l<-c(.5, 1, 2)
len_new_ls<-NULL; len_iqr_ls<-NULL; supp_new_ls<-NULL; dose_new_ls<-NULL
TG_new<-data.frame()
for (i in supp_l){
  for (n in dose_l){
    len_new<-mean(ToothGrowth[which(ToothGrowth$supp==i & ToothGrowth$dose==n), ]$len)
    len_iqr<-IQR(ToothGrowth[which(ToothGrowth$supp==i & ToothGrowth$dose==n), ]$len)
    len_new_ls<-c(len_new_ls, len_new)
    len_iqr_ls<-c(len_iqr_ls, len_iqr)
    supp_new_ls<-c(supp_new_ls, i)
    dose_new_ls<-c(dose_new_ls, n)
  }
}
TG_new<-data.frame(supp=supp_new_ls, dose=dose_new_ls, len_avg=len_new_ls, len_iqr=len_iqr_ls)
print(TG_new)
```

```
##      supp dose len_avg len_iqr
## 1    OJ  0.5   13.23   6.475
## 2    OJ  1.0   22.70   5.350
## 3    OJ  2.0   26.06   2.500
## 4    VC  0.5    7.98   4.950
## 5    VC  1.0   16.77   2.025
## 6    VC  2.0   26.14   5.425
```

From the TG_new, we could get the basic conclusion: 1. The tooth growth is related with the dose: The more dose takes, the longer the tooth grow; 2. Different supplies impact the tooth growth: People used the supply of OJ have the longer tooth growth. The plots below also shows the similar situation.

```
plot(ToothGrowth[which(ToothGrowth$supp=="OJ"), 1], ToothGrowth[which(ToothGrowth$supp=="OJ"), 3], pch=1, col="red",
     points(ToothGrowth[which(ToothGrowth$supp=="VC"), 1], ToothGrowth[which(ToothGrowth$supp=="VC"), 3], pch=1, col="blue"),
     legend(10, 2, c("OJ", "VC"), pch=c(0, 2), col=c("red", "blue"))
```



Compare tooth growth by supp and dose: From the histogram, we found that the histograms with the supp OJ and supp VC are not strictly symmetric. So, I use the t.test to compare the data. (I did not show these two histograms here due to the report length request.)

Use t test on the supp first. We found the confidence interval of OJ-VC is [1.4, 6.0] with 95% confidence level. This means using supp OJ, the tooth growth longer than supp VC. If we use the hypothesis test, the H0 means we assume supp OJ and supp VC have the same impact on the tooth growth. And the H1, alternative assumption, is supp OJ and supp VC have different impact on the tooth growth. And as usual, we set the Type I error level at .05. From the t.test result, the p value for the OJ-VC is .0025, which is smaller than .05. This means the tooth growth with supp OJ and VC have significant different impact, so we reject the H0.

```
t_OJ<-ToothGrowth[which(ToothGrowth$supp=="OJ"), ]$len; t_VC<-ToothGrowth[which(ToothGrowth$supp=="VC")$len]
t.test(t_OJ-t_VC, paired=FALSE, var.equal=FALSE)
```

```
##
## One Sample t-test
##
## data: t_OJ - t_VC
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 1.408659 5.991341
## sample estimates:
## mean of x
## 3.7
```

Use t test on the dose, and we compare the dose .5 vs dose 1 and dose1 vs dose2 separately. We found the confidence interval of dose.5-dose1 is [-11.87, -6.39] and dose1-dose2 is [-9.26, -3.47] with 95% confidence level. Both demonstrate more dose will have longer tooth growth. If we use the hypothesis test, the H0 means we

assume dose.5 and dose1 have the same impact on the tooth growth, and the H1, alternative assumption, is dose.5 and dose1 have different impact on it. We set the type I error level at .05. From the t.test result, the p value for the dose.5-dose1 is 1.225e-06. This means tooth growth with different doses has significant different impact, so we reject H0. So is the dose1-dose2. The p value is .0002 with .05 type I error. So, we reject H0. Both hypothesis tests show that different dose have different impact on the tooth growth.

```
dose_.5<-ToothGrowth[which(ToothGrowth$dose==.5), ]$len
dose_1<-ToothGrowth[which(ToothGrowth$dose==1), ]$len
dose_2<-ToothGrowth[which(ToothGrowth$dose==2), ]$len
t.test(dose_.5-dose_1, paired=FALSE, var.equal=FALSE)
```

```
##
## One Sample t-test
##
## data: dose_.5 - dose_1
## t = -6.9669, df = 19, p-value = 1.225e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -11.872879 -6.387121
## sample estimates:
## mean of x
## -9.13
```

```
t.test(dose_1-dose_2, paired=FALSE, var.equal=FALSE)
```

```
##
## One Sample t-test
##
## data: dose_1 - dose_2
## t = -4.6046, df = 19, p-value = 0.0001934
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -9.258186 -3.471814
## sample estimates:
## mean of x
## -6.365
```

Summary: Applying more dose with the same supp or choosing OJ with the same dose will make the tooth grow longer.