

# ToothGrowth Data Analysis

*Liang Dong*

*25 October 2014*

## Contents

|                                     |          |
|-------------------------------------|----------|
| <b>ToothGrowth Data Analysis</b>    | <b>1</b> |
| Basic Data Exploratory . . . . .    | 1        |
| Basic Summary of the Data . . . . . | 2        |
| Hypothesis test . . . . .           | 4        |
| <b>Conclusion</b>                   | <b>5</b> |

## ToothGrowth Data Analysis

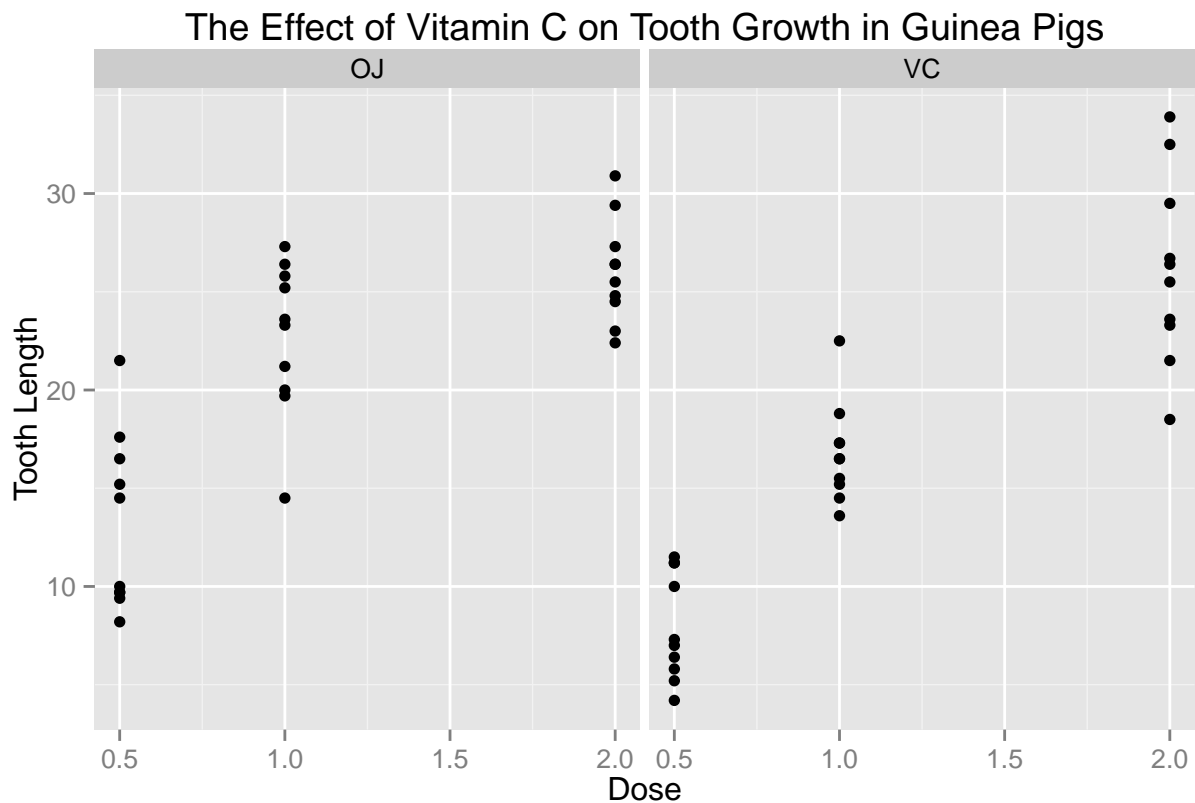
### Basic Data Exploratory

First load the library into the R.

```
library(ggplot2)
library(datasets)
```

Then use ggplot to plot the ToothGrowth data on different supp.

```
ggplot(data=ToothGrowth, aes(x=dose, y=len)) + facet_wrap(~supp) + geom_point() +
  xlab("Dose") + ylab("Tooth Length") +
  ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs")
```



## Basic Summary of the Data

The OJ and VC are separated and summarized by different dose level.

```
VC <- subset(ToothGrowth, supp == 'VC')
OJ <- subset(ToothGrowth, supp == 'OJ')
```

```
VC$supp <- factor(VC$supp)
OJ$supp <- factor(OJ$supp)
```

```
VC05 <- subset(VC, dose==0.5)
VC10 <- subset(VC, dose==1.0)
VC20 <- subset(VC, dose==2.0)
OJ05 <- subset(OJ, dose==0.5)
OJ10 <- subset(OJ, dose==1.0)
OJ20 <- subset(OJ, dose==2.0)
```

```
summary(VC05)
```

```
##      len      supp      dose
##  Min.   : 4.20   VC:10  Min.   :0.5
##  1st Qu.: 5.95           1st Qu.:0.5
##  Median : 7.15           Median :0.5
##  Mean   : 7.98           Mean   :0.5
```

```
## 3rd Qu.:10.90      3rd Qu.:0.5
## Max.      :11.50      Max.      :0.5
```

```
summary(OJ05)
```

```
##      len      supp      dose
## Min.   : 8.2    OJ:10  Min.   :0.5
## 1st Qu.: 9.7           1st Qu.:0.5
## Median :12.2           Median :0.5
## Mean   :13.2           Mean   :0.5
## 3rd Qu.:16.2           3rd Qu.:0.5
## Max.   :21.5           Max.   :0.5
```

```
summary(VC10)
```

```
##      len      supp      dose
## Min.   :13.6    VC:10  Min.   :1
## 1st Qu.:15.3           1st Qu.:1
## Median :16.5           Median :1
## Mean   :16.8           Mean   :1
## 3rd Qu.:17.3           3rd Qu.:1
## Max.   :22.5           Max.   :1
```

```
summary(OJ10)
```

```
##      len      supp      dose
## Min.   :14.5    OJ:10  Min.   :1
## 1st Qu.:20.3           1st Qu.:1
## Median :23.4           Median :1
## Mean   :22.7           Mean   :1
## 3rd Qu.:25.6           3rd Qu.:1
## Max.   :27.3           Max.   :1
```

```
summary(VC20)
```

```
##      len      supp      dose
## Min.   :18.5    VC:10  Min.   :2
## 1st Qu.:23.4           1st Qu.:2
## Median :25.9           Median :2
## Mean   :26.1           Mean   :2
## 3rd Qu.:28.8           3rd Qu.:2
## Max.   :33.9           Max.   :2
```

```
summary(OJ20)
```

```
##      len      supp      dose
## Min.   :22.4    OJ:10  Min.   :2
## 1st Qu.:24.6           1st Qu.:2
## Median :25.9           Median :2
## Mean   :26.1           Mean   :2
## 3rd Qu.:27.1           3rd Qu.:2
## Max.   :30.9           Max.   :2
```

## Hypothesis test

From the exploratory analysis and basic summary from first and second section, the mean length of 0.5 and 1.0 dose level from OJ is larger than VC. But on dose level 2.0, the mean length is equal, so it not obvious on which one is higher length.

Base on this initial analysis, suppose the  $H_0$  and  $H_a$  on the same dose level to be:

- $H_0$ : The length of using supplement VC is greater than OJ
- $H_a$ : The alternative hypothesis is length of using supplement VC is no more than method OJ

The same dose level are compared so only three group are compared. Assuming the significant level is 0.05.

### On dose level 0.5

```
t.test(OJ05$len, VC05$len, var.equal = FALSE, alternative = "greater", conf.level = 0.95)

##
## Welch Two Sample t-test
##
## data: OJ05$len and VC05$len
## t = 3.17, df = 14.97, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.346      Inf
## sample estimates:
## mean of x mean of y
##    13.23      7.98
```

The p-value is less than 0.05, so we **reject the null hypothesis**, which means the result is confirming that VC has NO greater impact than OJ on dose level 0.5.

### On dose level 1.0

```
t.test(OJ10$len, VC10$len, var.equal = FALSE, alternative = "greater", conf.level = 0.95)

##
## Welch Two Sample t-test
##
## data: OJ10$len and VC10$len
## t = 4.033, df = 15.36, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356      Inf
## sample estimates:
## mean of x mean of y
##    22.70    16.77
```

The p-value is less than 0.05 as well, so we **reject the null hypothesis** too, which also means the VC has NO greater impact than OJ on dose level 1.0.

On dose level 2.0

```
t.test(OJ20$len, VC20$len, var.equal = FALSE, alternative = "greater", conf.level = 0.95)
```

```
##  
## Welch Two Sample t-test  
##  
## data: OJ20$len and VC20$len  
## t = -0.0461, df = 14.04, p-value = 0.5181  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## -3.133 Inf  
## sample estimates:  
## mean of x mean of y  
## 26.06 26.14
```

This time p-value is significant larger than 0.05, so **null hypothesis is NOT rejected**, which confirming the VC has greater impact than OJ on dose level 2.0.

## Conclusion

The conclusion are:

- On dose level 0.5 and 1.0, supplement OJ is better than VC.
- On dose level 2.0, supplement OJ is not better than VC.