

# Project 1/2 - Statistic Inference

*ctakamiya*

*September 14, 2014*

## 1 - Loading the ToothGrowth data and performing some basic exploratory data analyses

```
data (ToothGrowth)
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
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
```

```
tail(ToothGrowth)
```

```
##      len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

## 2 - A basic summary of the data.

Average of length by type supplementar

```
summary(ToothGrowth[ToothGrowth$supp=="OJ",]$len)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       8.2    15.5    22.7    20.7    25.7    30.9
```

```
summary(ToothGrowth[ToothGrowth$supp=="VC",]$len)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.2    11.2    16.5    17.0    23.1    33.9
```

### 3 - Using confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

After initial exploratory data analysis, this data provide an intuition about the difference between OJ and VC averages in length. In another words,  $\mu_{OJ} > \mu_{VC}$  or  $\mu_{OJ} - \mu_{VC} > 0$ . This test is not paired data, because the observations are independent.

$H_0 : \mu_{OJ} - \mu_{VC} = 0$  - There is no difference in length averages.

$H_A : \mu_{OJ} - \mu_{VC} \neq 0$  - There is a difference in length averages.

```
t.test(ToothGrowth[ToothGrowth$supp=="OJ",]$len,
       ToothGrowth[ToothGrowth$supp=="VC",]$len,
       var.equal = FALSE)
```

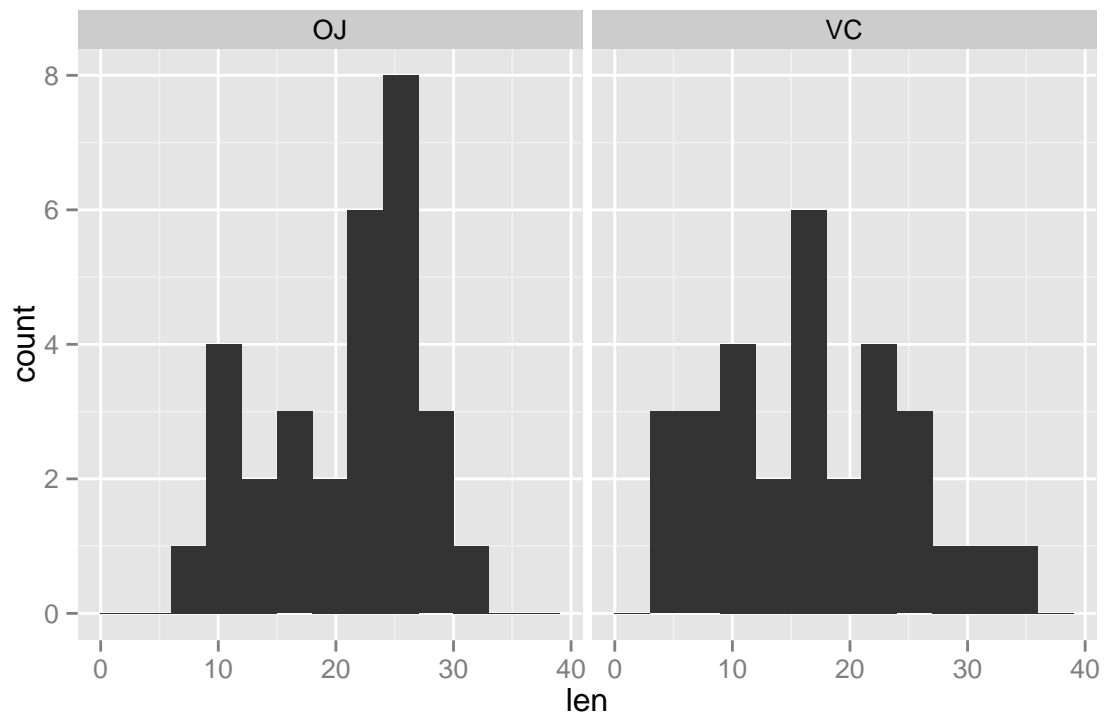
```
##
## Welch Two Sample t-test
##
## data:  ToothGrowth[ToothGrowth$supp == "OJ", ]$len and ToothGrowth[ToothGrowth$supp == "VC", ]$len
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean of x mean of y
##      20.66      16.96
```

### 4 Conclusions

The result in part 3 shows that there is no difference in length averages. Using Confidence Interval (95%) we can verify that 0 (zero) is in the interval. Or by significance  $\alpha = 0.05$  and p-value  $> 0.05$  (p-value=0.06063), we fail to reject  $H_0$ . In another word, this data do not provide convincing evidence that OJ promotes more growth of tooth length.

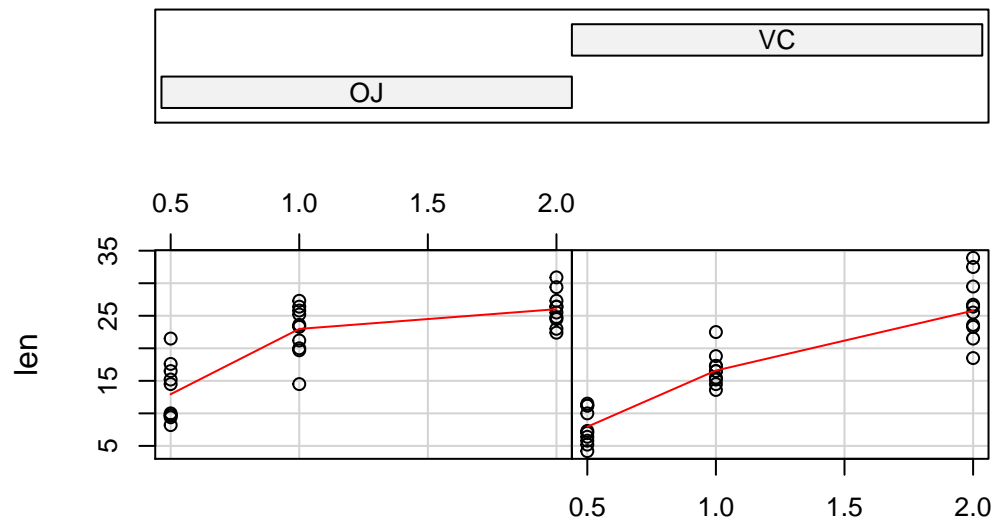
### Appendix

```
library(ggplot2)
qplot(len, data = ToothGrowth, facets=~supp, binwidth=3)
```



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

Given : supp



ToothGrowth data: length vs dose, given type of supplement