# Part 2 of Statistical Inference Project

***Allen S.C. Tang***

***2015年3月22日***

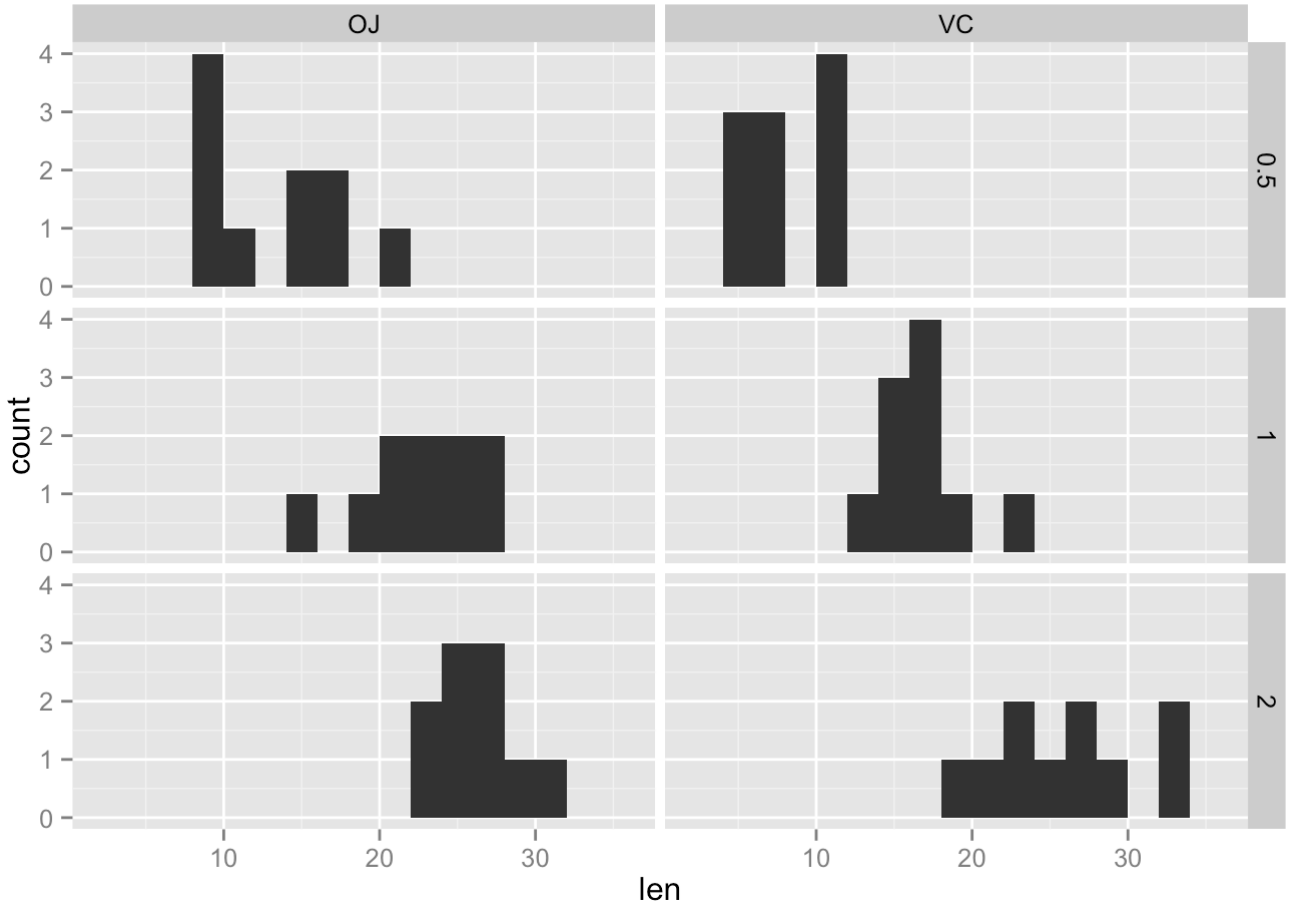
## Basic Exploratory Data Analysis

To undersatnd the basic status of ToothGrowth dataset, the exploratory data analysis is conducted as below:

```
library(ggplot2)
data(ToothGrowth)
tooth<-ToothGrowth
str(tooth)
```

```
## '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 ...
```

```
##unique(tooth$dose)
g1<-ggplot(tooth,aes(x=len))+geom_histogram(binwidth=2)+facet_grid(dose~supp)
plot(g1)
```




## Basic Summary of Data

The ToothGrowth dataset contains 60 observations of 3 variables: len, supp and dose. The supp variable contains

2 factors of “VC” and “OJ”, and the dose variable has values of 0.5, 1.0 and 2.0. I separate the original dataset by supp and dose variables and plot the 2x3 panel histogram to get a clearer view of the data. According to the panel historgram, the effect of 2 supp varaibles seem the same, and the increase of dose variable means more effective.

# Hypothesis Tests

First test is to determine if the 2 factors “VC” and “OJ” in supp variable are equal. I divide the original dataset into 2 subsets by supp “VC” and “OJ”, then compare the means of 2 subsets to get the result. ####H0: The means of len variables in 2 subsets are equal. ####Ha: The means of len variables in 2 subsets are unequal. ####The confidence interval is 95%.

```
tooth$supp<-as.character(tooth$supp)
VCset<-subset(tooth,tooth$supp=="VC")
OJset<-subset(tooth,tooth$supp=="OJ")
t.test(x=VCset$len,y=OJset$len,alternative=c("two.sided"),mu=0,var.equal=T,conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  VCset$len and OJset$len
## t = -1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5670064  0.1670064
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Based on the t test result, the t value is in the confidence interval, which means that it’s failed to reject H0, and 2 means from 2 subsets are equal.

Second test is to determine if the dose is increased, the effect to tooth grwoth is increased too. I divide the original dataset into 3 subsets by dose 0.5, 1.0 and 2.0, and compare the means of 3 subsets to get the result. ####H0: The mean of the subset with more dose is larger than one with less dose. ####Ha: The mean of the subset with more dose is not larger than one with less dose. ####The confidence interval is 95%.

```
set05<-subset(tooth,tooth$dose==0.5)
set10<-subset(tooth,tooth$dose==1.0)
set20<-subset(tooth,tooth$dose==2.0)
t.test(x=set10$len,y=set05$len,alternative=c("less"),mu=0,var.equal=T,conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  set10$len and set05$len
## t = 6.4766, df = 38, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 11.50666
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```
t.test(x=set20$len,y=set10$len,alternative=c("less"),mu=0,var.equal=T,conf.level=0.95)
```

```
##
##  Two Sample t-test
##
## data:  set20$len and set10$len
## t = 4.9005, df = 38, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 8.554804
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Based on the above t test results, the 2 t values are both in the confidence intervals, which means that both tests are failed to reject H0. This says the mean of dose 2.0 subset is greater than the one of dose 1.0 subset, and the mean of dose 1.0 is greater than the one of dose 0.5.

# Assumptions to Get The Conclusions

1. The variances of populations of comparing pair of subsets are equal.
2. The observations are all i.i.d.
3. The popluation of every subset follows normal distribution.

# Conclusions

Under previous assumptions, the reading of results from hypothesis tests are: 1. The means of the OJ and VC subsets are equal, which represents that the effects to tooth growth by using OJ and VC are equal too. 2. The means are increasing when dose increases, which represents that the more dose given to the subject, the more tooth growth can be expected.