Statistical Inference Course Project Part 2

Shun-Wen Chang

October 24, 2015

Overview:

This project analyzes the ToothGrowth data in the R datasets package and uses confidence intervals and hypothesis tests to compare tooth growth by supp and dose. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Load Data:

Load data and necessary packages.

```
library(ggplot2)
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
```

Summary of the Data:

```
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
```

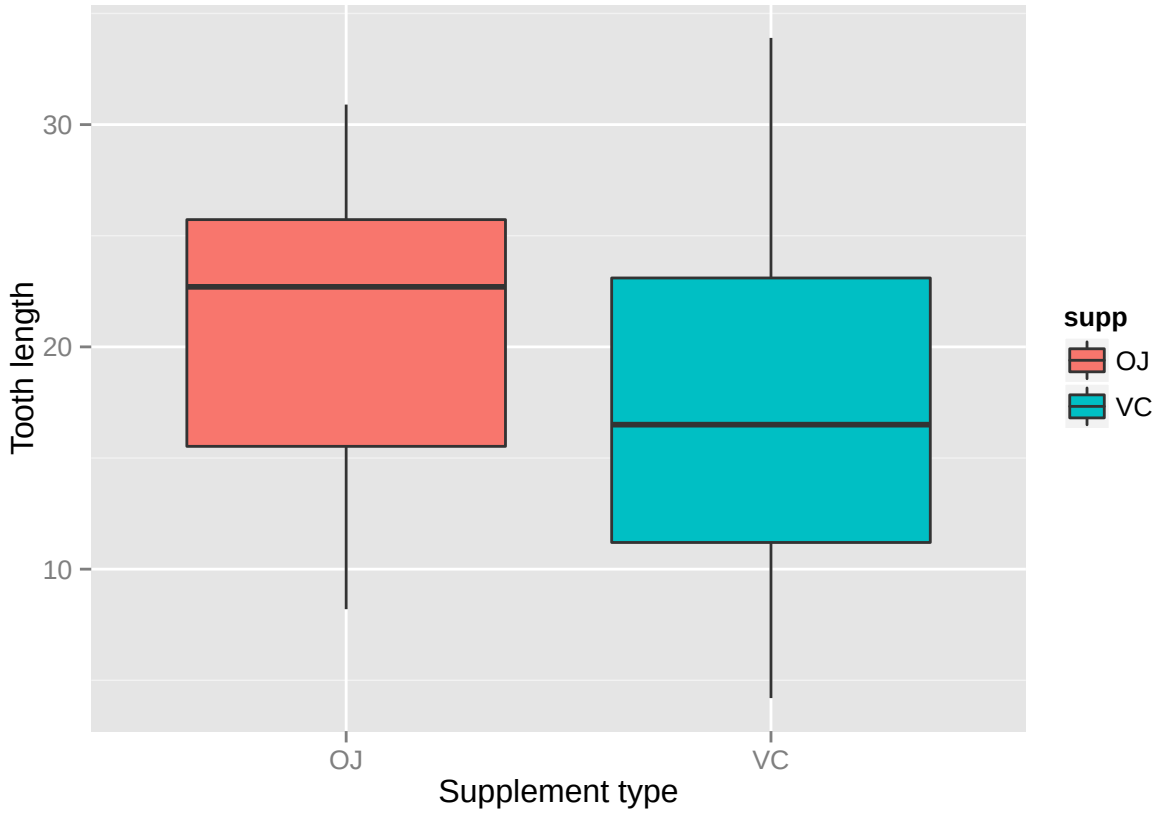
Exploratory Analysis:

Divide the tooth length data by delivering method and create box plot for viewing.

```
Toothlength_deliver <- split(ToothGrowth$len,ToothGrowth$supp)
sapply(Toothlength_deliver,mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

```
ggplot(aes(x=supp,y=len),data=ToothGrowth) + geom_boxplot(aes(fill=supp))+ xlab("Supplement type")+ylab
```



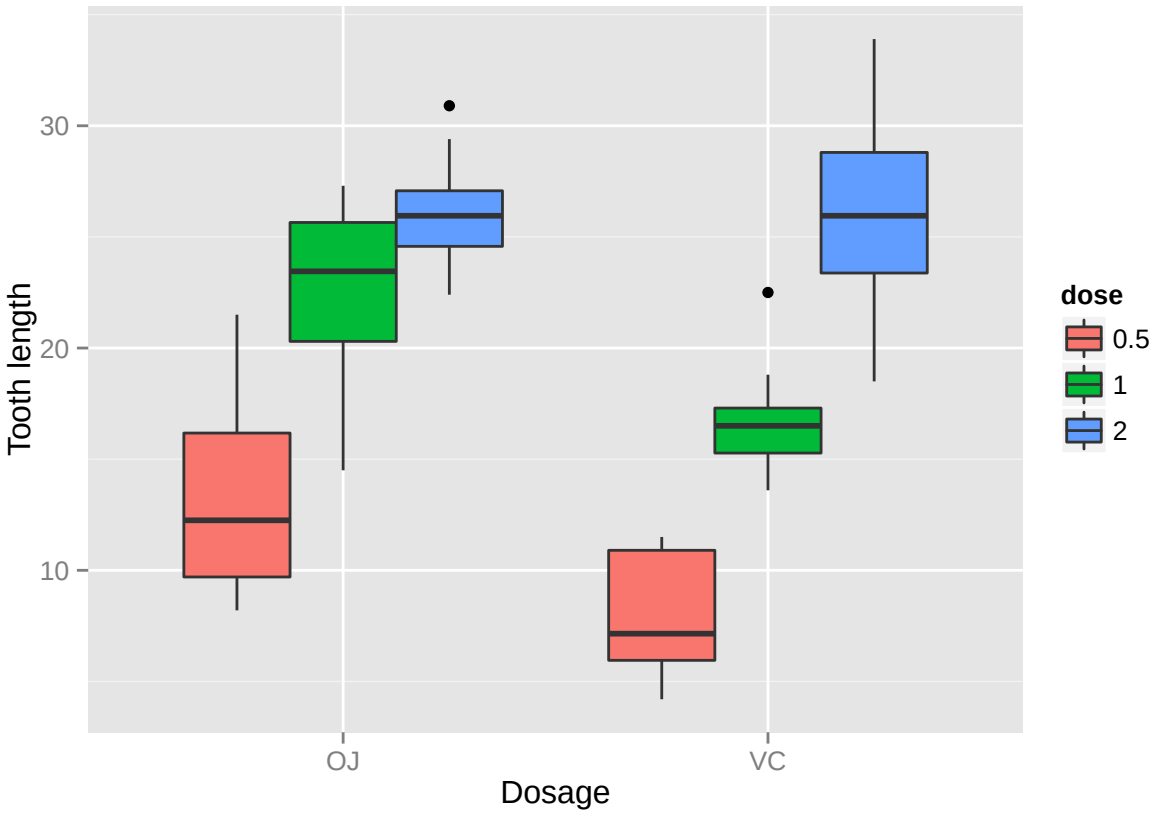
It may seem like with method OJ the tooth length is on averagy longer. However, we need to do further check. . .

Now we check the tooth length with different dose:

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
Toothlength_dose <- split(ToothGrowth$len,ToothGrowth$dose)
sapply(Toothlength_dose,mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
ggplot(aes(x=supp,y=len),data=ToothGrowth) + geom_boxplot(aes(fill=dose))+ xlab("Dosage")+ylab("Tooth l
```



It may seem like with higher dose the tooth length is on averagy longer. However, we need to do further check...

Use Hypothesis Test to Compare data:

Assumption: significance level is 5%

Test if the tooth growth length depends on the vitamin delivery methods.

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

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[supp == "VC"]
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

p-value = 0.06063 is over significance level 5%. We cannot reject null hypothesis. Therefore we conclude that the delivery methods has no effect on the tooth growth.

Test if the tooth growth length depends on the vitamin dosage.

```
t.test(len[dose==2], len[dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len[dose == 2] and len[dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

p-value is almost zero, less than 5%. We can therefore reject null hypothesis. We then conclude that the dosage cause positive effect on the tooth growth, which means that higher dosage will lead to longer tooth.

My conlusion is that vitamin delivery methods doesn't affect the tooth growth, while higher dosage leads to longer tooth growth.