Untitled

Roland Feng

Tuesday, July 21, 2015

Course Project for Statistical Inference Part 2

By Roland Feng

The data is about the growth of teeth. There are 60 rows in the ToothGrowth and there are three colomns corresponding to the length of teeth, dilivery methods(orange juice or ascorbic acid) and dose levels of Vitamin C(0.5,1,2mg).

Load the ToothGrowth data and perform some basic exploratory data analyses

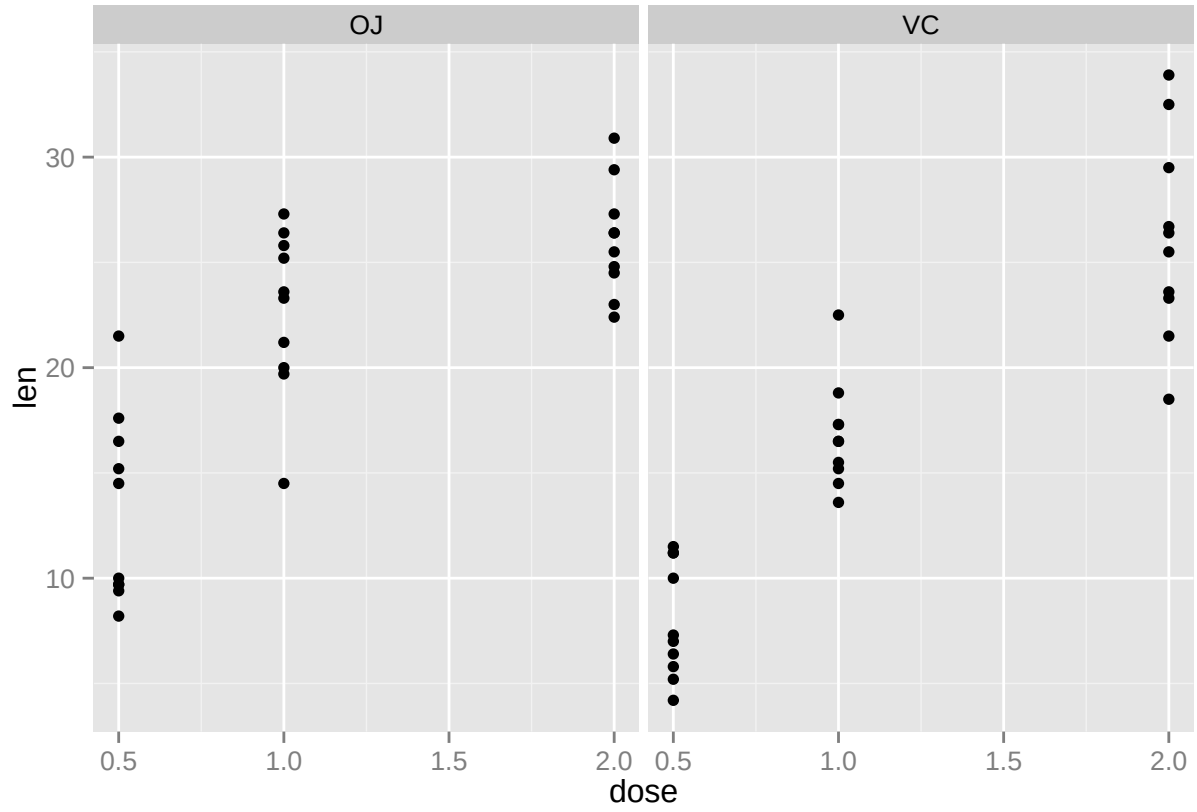
Firtst we load the data into R.

```
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 ...
```

I decide to explore the relationship between the length and dose. We seperate the data by supp and draw the figures.

```
library(ggplot2)
qplot(dose, len, data=ToothGrowth,facets=.~supp)
```



It is obvious from the picture that the length of tooth has a positive relationship with the dose.

Provide a basic summary of the data.

We transform the "dose" into a factor show the summary.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

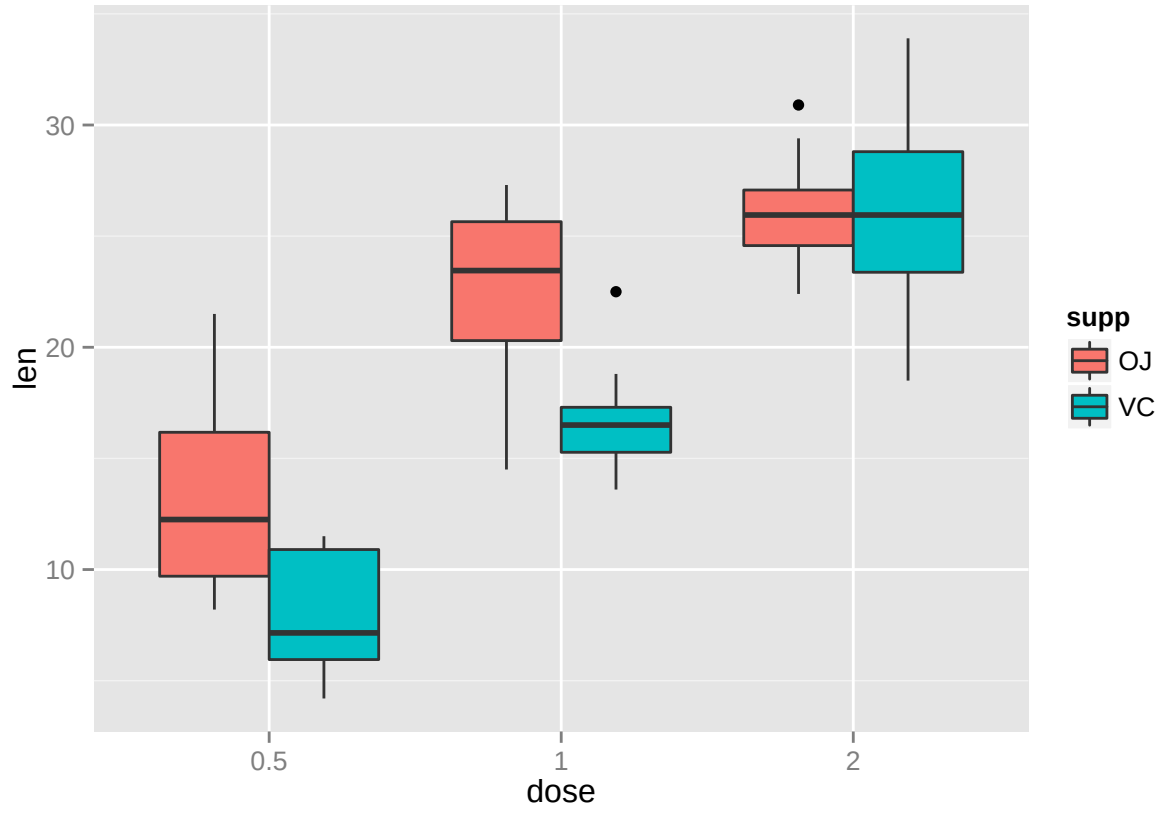
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
table(ToothGrowth$dose,ToothGrowth$supp)
```

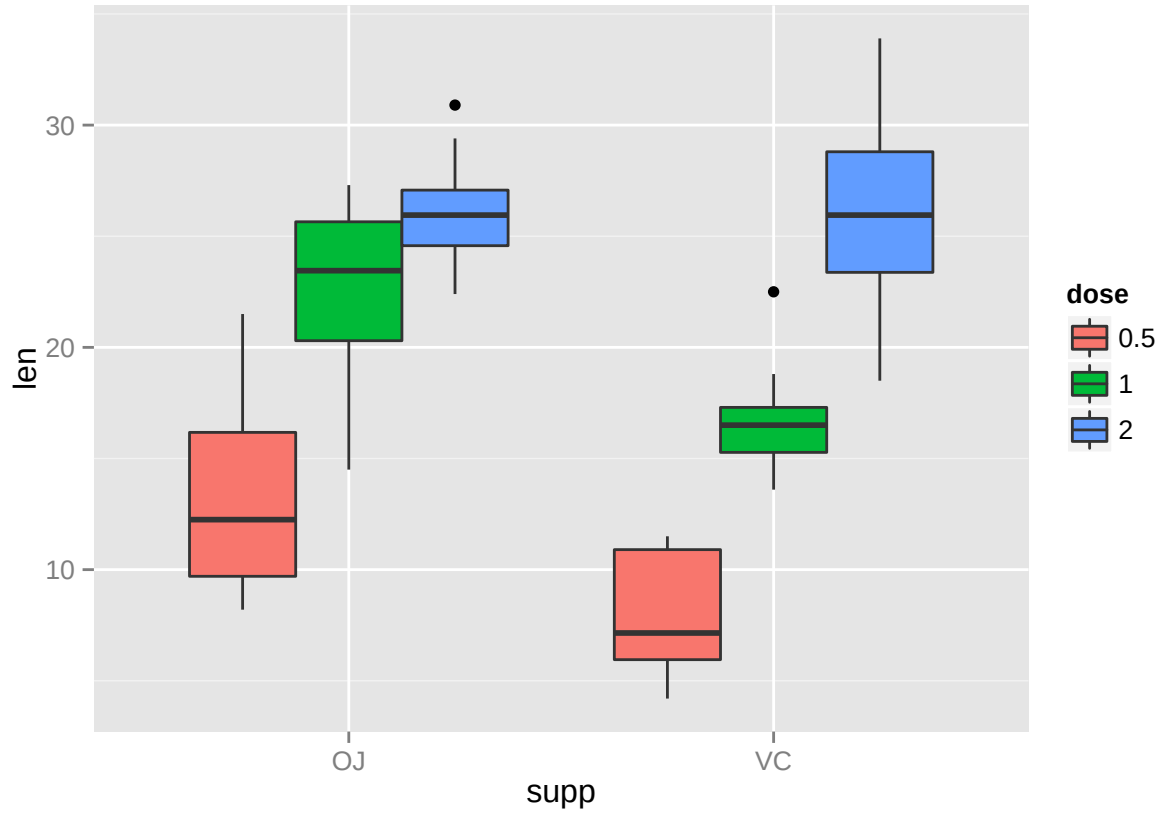
```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

Then we draw some box figure related to the data.

```
ggplot(aes(x=dose,y=len),data=ToothGrowth)+geom_boxplot(aes(fill=supp))
```




```
ggplot(aes(x=supp,y=len),data=ToothGrowth)+geom_boxplot(aes(fill=dose))
```



Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

We will compare the len from the two type of delivery methods. We set the null hypothesis as the delivery has no effect on the length of teeth.

```
lapply(split(ToothGrowth$len,ToothGrowth$supp),var)
```

```
## $OJ
## [1] 43.63344
##
## $VC
## [1] 68.32723
```

So the variance could not be regarded as the same.

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

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$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 is 0.061, and the confidence interval is (-0.171,7.571), which contains 0. So we can not reject the null hypothesis.

Next we compare the len from the dose. The null hypothesis is that the dose does not effect the len.

```
lapply(split(ToothGrowth$len,ToothGrowth$dose),var)
```

```
## $`0.5`
## [1] 20.24787
##
## $`1`
## [1] 19.49608
##
## $`2`
## [1] 14.24421
```

By the same reason, we will not regard the variances as the same.

```
t.test(ToothGrowth$len[ToothGrowth$dose=="0.5"],ToothGrowth$len[ToothGrowth$dose=="1"],paired=FALSE,var
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == "0.5"] and ToothGrowth$len[ToothGrowth$dose == "1"]
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(ToothGrowth$len[ToothGrowth$dose=="0.5"],ToothGrowth$len[ToothGrowth$dose=="2"],paired=FALSE,var
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == "0.5"] and ToothGrowth$len[ToothGrowth$dose == "2"]
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
t.test(ToothGrowth$len[ToothGrowth$dose=="1"],ToothGrowth$len[ToothGrowth$dose=="2"],paired=FALSE,var.e
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == "1"] and ToothGrowth$len[ToothGrowth$dose == "2"]
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

The p-values are 1.268e-07, 4.398e-14, 1.906e-05. There are all less than 0.05. The confidence intervals are (-11.983781 -6.276219), (-18.15617 -12.83383), (-8.996481 -3.733519). All the confidence intervals do not contain 0. According to all, we can reject the null hypothesis.

conclusions and the assumptions

Conclusion

1. There isn't enough evidence to support the idea that delivery methods have effect on the length of teeth.
2. The dose level has a positive effect on the length of teeth.

Assumption

1. All the measurements are reasonable. The samples is representive and selected randomly.
2. We set all the variances to be different, which is more objective.