Statistical Inference Course Project - Part 2

Fu-Ching Yang

Monday, October 03, 2016

Overview

In this project, we will first load and glance the ToothGrowth data. Then, we will performance hypothesis test on the impact to tooth growth in different aspects. Finally, we will draw conclusion and suggestions.

Load data

The specified data is ToothGrowth. Load it with data().

```
data(ToothGrowth)
```

Basic summary of the data

Use head() to glance the data.

```
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
```

With str(), we know ToothGrowth is a data frame with 60 data in 3 columns: len, supp and dose. len and dose are numerical data, while supp is a categorical data having 2 categories: OJ and VC.

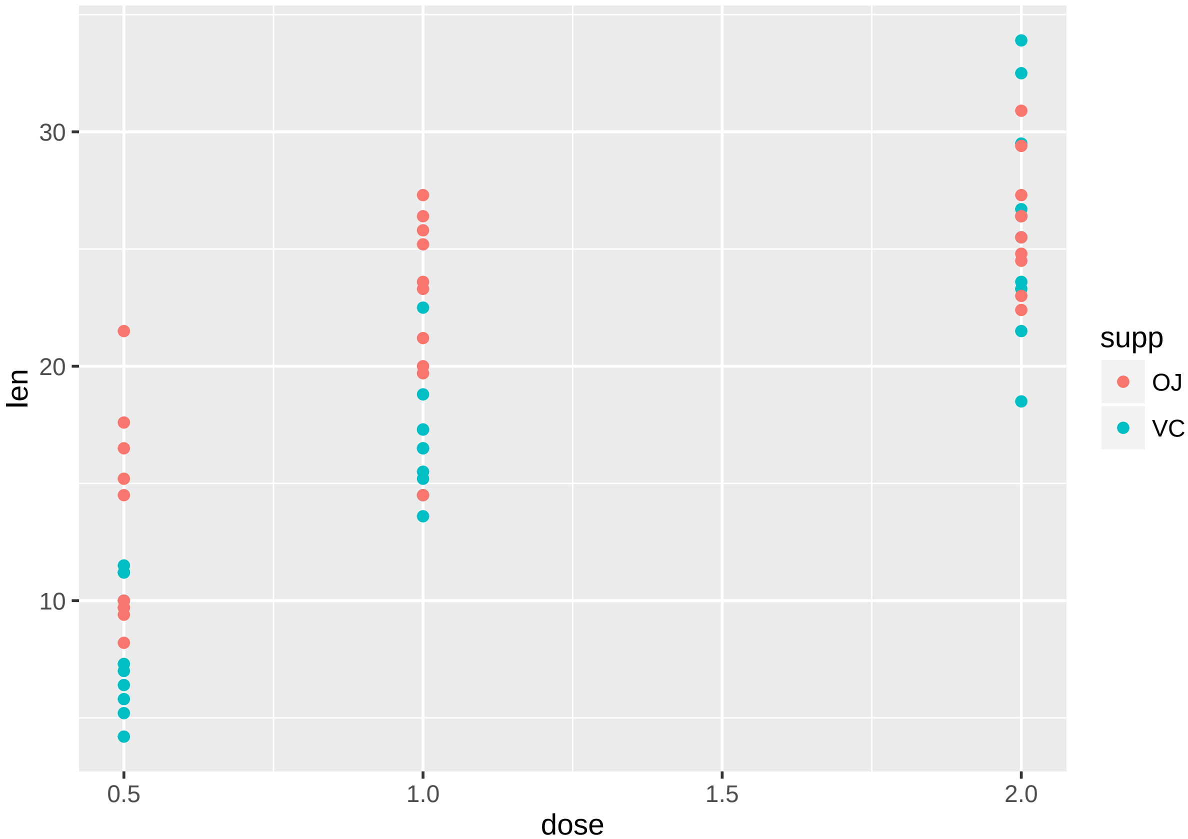
```
str(ToothGrowth)
```

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

The data appears to document the tooth growth length with different dosage, in two groups (OJ and VC).

To have a feeling of the data relationship in 3 variables, we perform scatter plot with supp in different color.

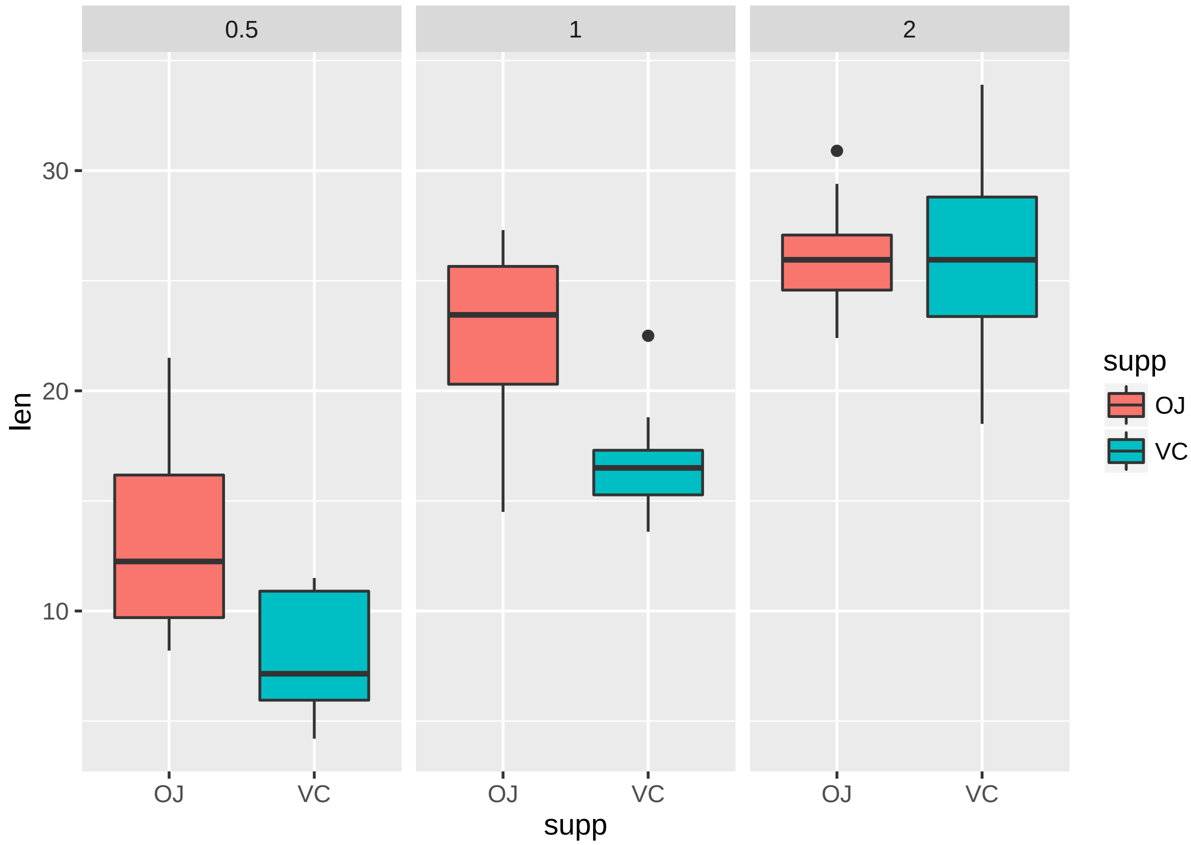
```
ggplot(data=ToothGrowth, aes(x=dose, y=len)) + geom_point(aes(colour=supp))
```



It looks like the dosage has only 3 types: 0.5, 1.0 and 2.0. And, the tooth growth of OJ seems to be greater than that of VC.

To better compare their variance, we change to boxplot.

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



With boxplot, it is clear that as the dosage increases the tooth growth increases, regardless of OJ and VC.

As for OJ and VC comparison, OJ tends to have more tooth growth than VC at 0.5 and 1 dosage, but have same average tooth growth at 2 dosage.

Hypothesis tests

Hypothesis for VC vs OJ

To understand whether the tooth growth of OJ versus VC is significant in different dosage, we will perform t.test(). Please note that, here we assume the population distribution is Normal and the data are randomly sampled.

Let's divide the data by their dosage, so we can compare OJ and VC in each dosage.

```
half_dose <- ToothGrowth[ToothGrowth$dose==0.5,]
one_dose <- ToothGrowth[ToothGrowth$dose==1,]
two_dose <- ToothGrowth[ToothGrowth$dose==2,]
```

VC vs OJ at 0.5 dosage

We use t.test() to test the tooth growth.

The null hypothesis: no difference between VC and OJ.

The alternative hypothesis: OJ has more tooth growth than VC.

(Note that, paired must set to FALSE, since we don't know if the test subject can be paired. var.equal must be FALSE either, since their variance is different.)

```
t.test(half_dose[half_dose$supp=="OJ",]$len, half_dose[half_dose$supp=="VC",]$len,
       alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  half_dose[half_dose$supp == "OJ", ]$len and half_dose[half_dose$supp == "VC", ]$len
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The results show that the p-Value (0.32%) is much lower than the specified alpha (5%). This means we need to reject the null hypothesis. In other words, OJ has more tooth growth than VC.

VC vs OJ at 1 dosage

Now, we look at the test at 1 dosage.

```
t.test(one_dose[one_dose$supp=="OJ",]$len, one_dose[one_dose$supp=="VC",]$len,
       alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  one_dose[one_dose$supp == "OJ", ]$len and one_dose[one_dose$supp == "VC", ]$len
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The results show that the p-Value (0.05%) is still much lower than the specified alpha (5%).

OJ still has more tooth growth than VC in 1 dosage.

VC vs OJ at 2 dosage

Now, we look at the test at 2 dosage.

```
t.test(two_dose[two_dose$supp=="OJ",]$len, two_dose[two_dose$supp=="VC",]$len,
       alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  two_dose[two_dose$supp == "OJ", ]$len and two_dose[two_dose$supp == "VC", ]$len
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Unsurprisingly, the results show that the p-Value (51%) is not lower than the specified alpha (5%). This indicates that OJ has the same tooth growth as VC in 2 dosage.

Hypothesis for dosage

Another interesting fact we want to know is, whether the dosage results in tooth growth, regardless of OJ and VC.

We use the divided data (half_dose, one_dose, two_dose) for this analysis.

0.5 vs 1 dosage

We look at the dosage comparison of 0.5 and 1.

The null hypothesis: no difference between 0.5 dosage and 1 dosage.

The alternative hypothesis: 1 dosage has more tooth growth than 0.5 dosage.

```r
t.test(one_dose$len, half_dose$len, alternative = "greater",
       paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  one_dose$len and half_dose$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

The results show that the p-Value is extremely lower than the specified alpha (5%). 1 dosage definitely has better tooth growth than 0.5 dosage.

1 vs 2 dosage

We look at the dosage comparison of 1 and 2.

The null hypothesis: no difference between 1 dosage and 2 dosage.

The alternative hypothesis: 2 dosage has more tooth growth than 1 dosage.

```r
t.test(two_dose$len, one_dose$len, alternative = "greater",
       paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  two_dose$len and one_dose$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The results show that the p-Value is still extremely lower than the specified alpha (5%). 2 dosage definitely has better tooth growth than 1 dosage.

Conclusion

Under the assumption that the population distribution is Normal and the data is random sampled, we perform t test to confirm facts of tooth growth significance.

From the hypothesis test result, we can confirm that OJ helps tooth growth better than VC at dosage of 0.5 and 1, but not at dosage of 2.

And, the dosage definitely helps tooth growth.

So, in general, more dosage can result in tooth growth. However, if the dosage is 2.0, then OJ or VC doesn't really matter. If dosage is 0.5 or 1, OJ is suggested to have better tooth growth.