ToothGrowth Data Analysis

X. Zeng

August 18, 2015

Introduction

In this report, we are going to analyze the ToothGrowth data in the R datasets package. 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).

Exploratory data analysis

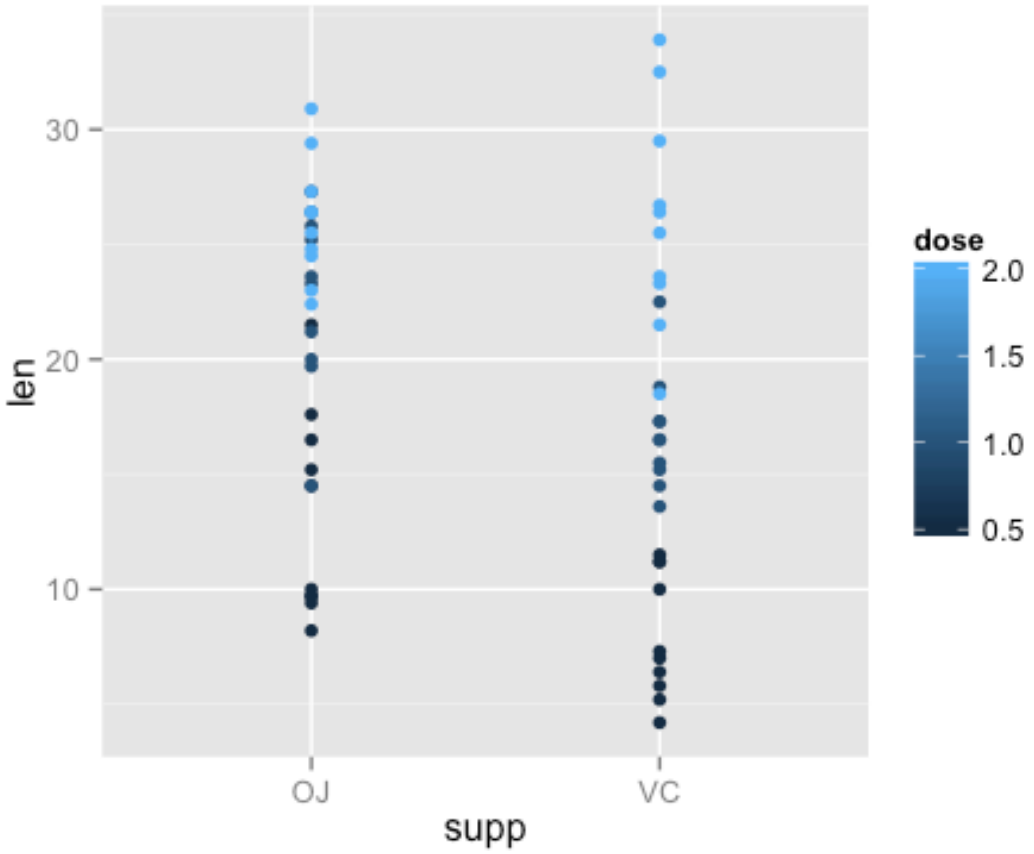
First, we need to load the data and take a look at the first three lines of the data.

```
library(ggplot2)
library(datasets)
head(ToothGrowth, 3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

Then, we can plot the tooth len with the supplies as independent variable, filled with the dose variable.

```
qplot(x=supp, y=len, color=dose, data=ToothGrowth)
```



It seems there are some differences for the tooth length among difference conditions, such as VC vs OJ, and different dose levels (0.5, 1.5, 2.0). However, it's hard to tell if the diffrences are real. So we need to do hypothesis tests to see if there are differences for the tooth lenght at different conditions. For now, we need summarise the data first.

```
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
```

Hypothesis tests

First, let's look at the effect of Vitamin C to see if there are diffrences between the means of tooth length treated by Vitamin C (VC) and Orange Juice (OJ). So we need to do a T-test.

```
t.test(ToothGrowth$len~ToothGrowth$supp,
       paired=FALSE, var.equal=FALSE, conf.level=0.95)

##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len by ToothGrowth$supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

From the results, we can see the p-value is larger than 0.05, which means we cannot reject the hypothesis that the means of tooth length treated by VC and OJ are equal at 95% confidence level.

Now, let's check the effect of different dose levels (0.5, 1.0, 2.0). Because there are 3 groups now, we need the one-way ANOVA.

```
oneway.test(ToothGrowth$len~ToothGrowth$dose, var.equal = FALSE)

##
##  One-way analysis of means (not assuming equal variances)
##
## data:  ToothGrowth$len and ToothGrowth$dose
## F = 68.401, num df = 2.000, denom df = 37.743, p-value = 2.812e-13
```

From the results, we can see the p-value is very small, so we can reject the null hypotheise, and conclude that there are differences for the tooth length among different dose levels.

Conclusion

For the above analysis, we cannot say there are differences between the tooth length treated by VC and OJ. However, the dose levels do affect the tooth length.

The above hypothesis analyses are based on the asummption that the variaces of tooth length are different for different groups, and the samples are independent with each other.

Appendix: R code

```
# This is for Statistical Inference Projcet2

# load libraries and data
library(ggplot2)
library(datasets)
head(ToothGrowth, 3)
```

```r
# Exploratory data analysis
qplot(x=supp, y=len, color=dose, data=ToothGrowth)

# provide summary
summary(ToothGrowth)

# statistical inference: effect of VS and OJ
t.test(ToothGrowth$len~ToothGrowth$supp,
       paired=FALSE, var.equal=FALSE, conf.level=0.95)

# statistical inference: effect of dose
oneway.test(ToothGrowth$len~ToothGrowth$dose, var.equal = FALSE)
```