ToothGrowth Analysis

zhenyue zhu

April 24, 2016

The purpose of this project is to analyze the ToothGrowth data in the R datasets package. The data includes the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Load the ToothGrowth data and perform some basic exploratory data analyses.

```r
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 ...
```

```r
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
```

```r
par(mfrow = c(1,2))
p1 <- ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill = factor(dose)))
p1 + geom_boxplot() + guides(fill=FALSE) + facet_grid(. ~ supp)
```
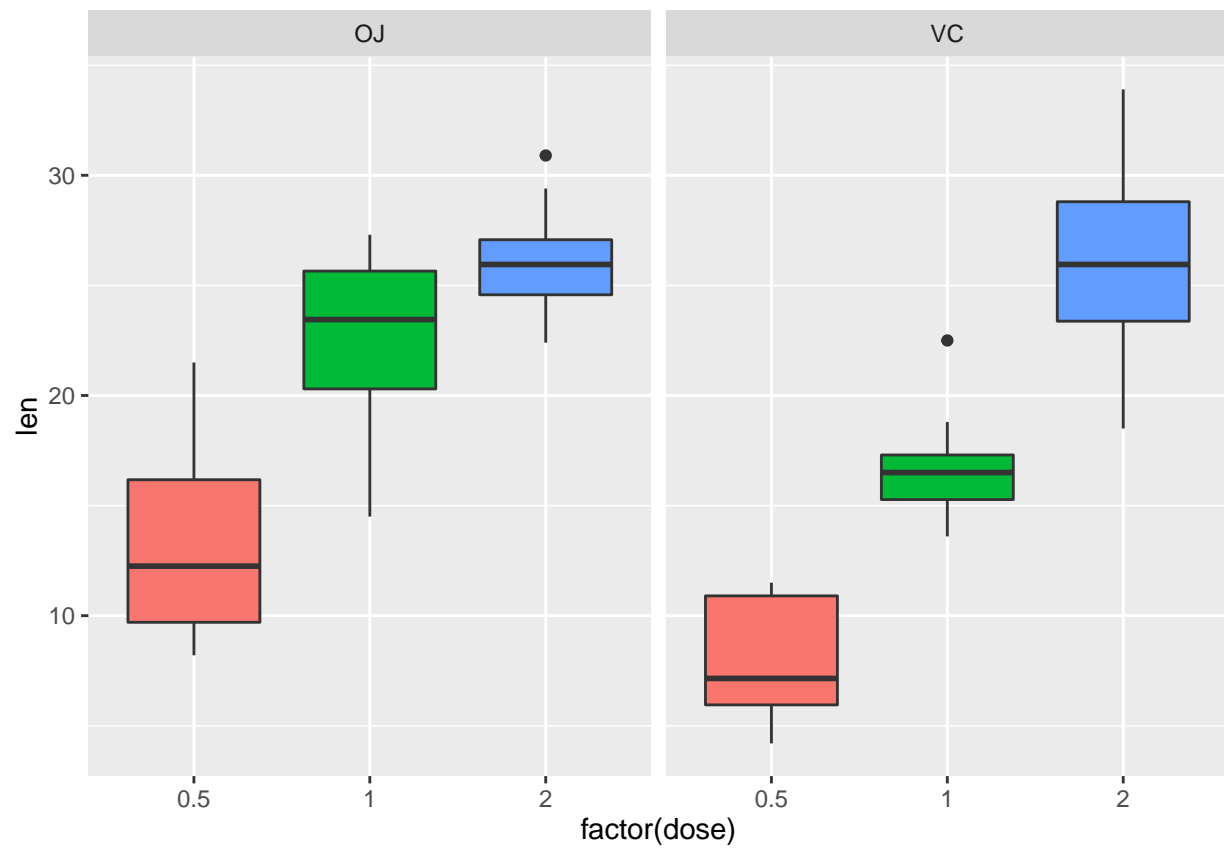
```
p2 <- ggplot(ToothGrowth, aes(x = factor(supp), y = len, fill = factor(supp)))
p2 + geom_boxplot() + guides(fill=FALSE) + facet_grid(. ~ dose)
```
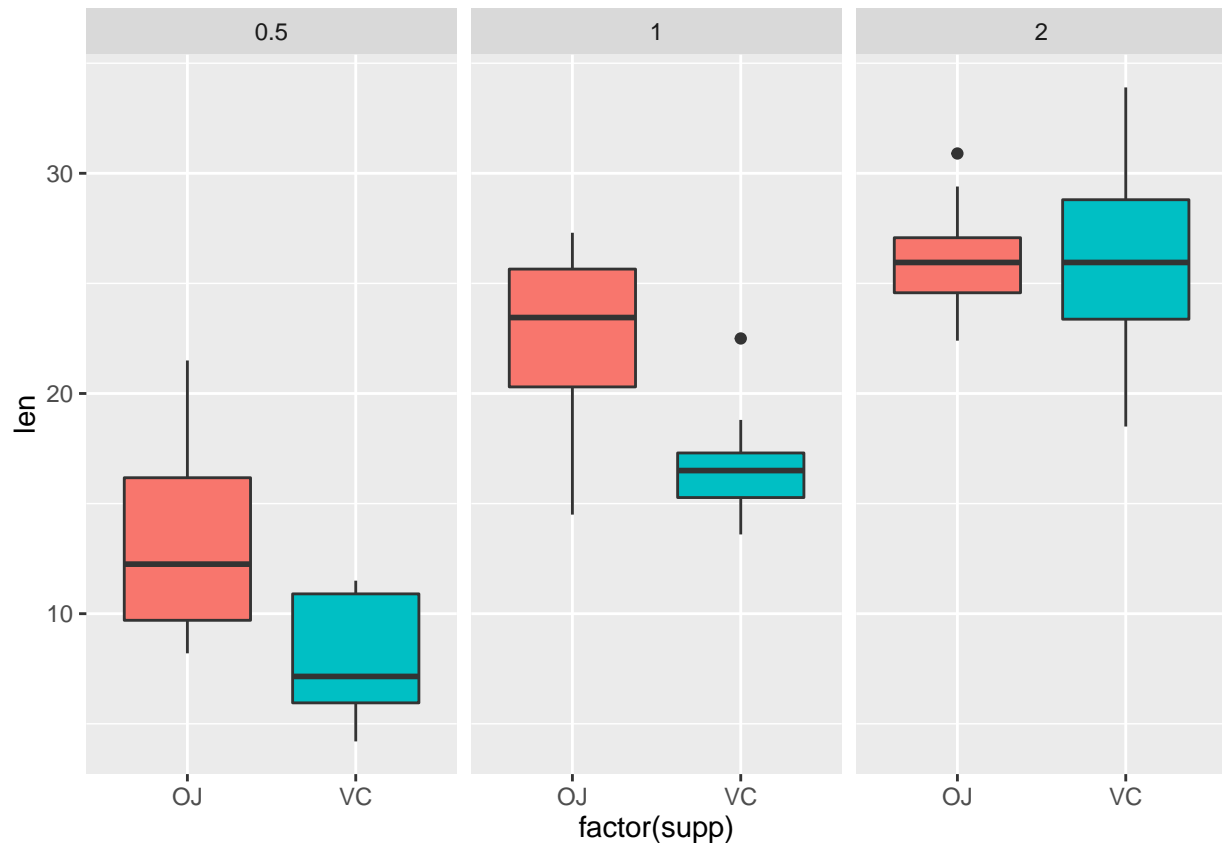
After plot the data, we see that for dose versus length, there is always an increasing trend, i.e. more dose, the tooth length will increase faster. But it looks like when a small dose is applied, the dose of OJ(orange juice) will make the teeth grow faster. At dose 2.0, it looks like both supplements has the similar effect on the tooth growth.

Statistical inferences

The tooth growth was compared by supplement at different dose, under the null hypothesis that each supplement has the same effect at the same dose for tooth growth. Then we divided the data sets into three doses groups. Then perform a t-test comparing length and supplements within each group.

```r
df0.5 <- subset(ToothGrowth, dose == 0.5)
df1.0 <- subset(ToothGrowth, dose == 1.0)
df2.0 <- subset(ToothGrowth, dose == 2.0)

test0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = df0.5)
test0.5$p.value
```

```
## [1] 0.006358607
```

```r
test0.5$conf
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
test1.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = df1.0)
test1.0$p.value
```

```
## [1] 0.001038376
```

```
test1.0$conf
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

For group dose is 0.5 or 1.0, we see that the p-value is very small (0.0064 and 0.0010 for dose 0.5 and 1.0). Confidence intervals are both positive. Indicating that the mean between each supplement is significant.

```
test2.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = df2.0)
test2.0$p.value
```

```
## [1] 0.9638516
```

```
test2.0$conf
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

However, for group with dose=2.0, the p-value is very high 0.96 and confidence interval includes negative. It indicate that for dose=2.0 group, different supplement has negligible effect on the tooth growth.

Assumptions.

I based the above inferences on the facts that:

- only orange juice and VC has effect on the tooth growth, without any other factors.
- 60 different samples are unrelated amd different samples has different variance. That's why I use paired=False and var.equal=False.
- The suject has identical other factors.

Conclusion.

Different supplements (VC and OJ) has different effects on tooth growth at lower dose according to t-test. At dose=0.5 and 1.0, OJ has stronger effect on the teeth growth. While at dose=2.0, there is no significant difference in tooth length comparing two supplements.