Statistical Inference Course Project

Composed by: Akihiro Hayashi (21/03/2015)

Summary

We reject that type of supplement doesn't affect tooth growth. And we also reject that dose deosn't affect tooth growth.

Outline of this project:

1. Summarize Dataset
2. Exploratory Data Analysis
3. Assumption
4. Test

Summarize Dataset

```
data(ToothGrowth)
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
```

```
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 ...
```

```
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
```

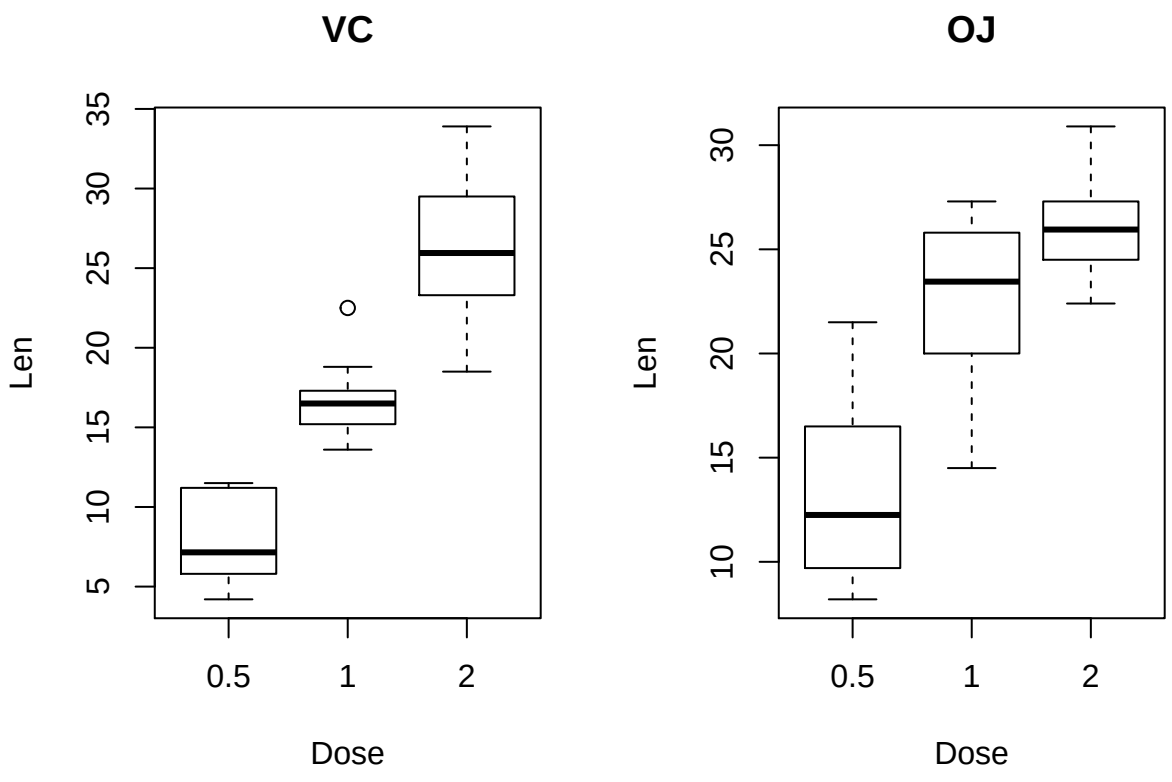
```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
df <- as.data.frame(group_by(ToothGrowth, supp, dose))
VC <- df[1:30, c(1, 3)]
OJ <- df[31:60, c(1, 3)]
```

Exploratory Data Analysis

```r
par(mfrow = c(1, 2))
boxplot(len ~ dose, data = VC, xlab = "Dose", ylab = "Len", main = "VC")
boxplot(len ~ dose, data = OJ, xlab = "Dose", ylab = "Len", main = "OJ")
```



Assumption

Paired Sample: From the description of this dataset, we know it's a paired sample with different means and variances when they are grouped by supp or dose.

Test

We use test to do our analysis (not confidence interval).

by Supp (H0: Supplement type doesn't affect tooth growth)

Rejection Rule: If supplement type doesn't affect tooth growth, p-value will be lower than significant level (.05).

```
pv <- round(t.test(len ~ supp, data = ToothGrowth, paired = TRUE, var.equal = FALSE)$p.value, 3)
```

Our p-value is 0.003 (lower than 0.05), so we reject H0. Supplement type will affect tooth growth.

by Dose (H0: Dose doesn't affect tooth growth)

We have to control supp variable. And as requirement, we cannot use ANOVA or regression, so we have to compare them 2 by 2.

```
vc0510 <- round(t.test(len ~ dose, data = VC[1:20, ], paired = TRUE, var.equal = FALSE)$p.value, 4)
vc0520 <- round(t.test(len ~ dose, data = VC[-(11:20), ], paired = TRUE, var.equal = FALSE)$p.value, 4)
vc1020 <- round(t.test(len ~ dose, data = VC[11:30, ], paired = TRUE, var.equal = FALSE)$p.value, 4)
oj0510 <- round(t.test(len ~ dose, data = OJ[1:20, ], paired = TRUE, var.equal = FALSE)$p.value, 4)
oj0520 <- round(t.test(len ~ dose, data = OJ[-(11:20), ], paired = TRUE, var.equal = FALSE)$p.value, 4)
oj1020 <- round(t.test(len ~ dose, data = OJ[11:30, ], paired = TRUE, var.equal = FALSE)$p.value, 4)
p.value <- cbind(vc0510, vc0520, vc1020, oj0510, oj0520, oj1020)
colnames(p.value) = c("VC0510", "VC0520", "VC1020", "OJ0510", "OJ0520", "OJ1020")
test <- (p.value < 0.05)
rownames(test) <- "Result"
test
```

```
##        VC0510 VC0520 VC1020 OJ0510 OJ0520 OJ1020
## Result   TRUE   TRUE   TRUE   TRUE   TRUE  FALSE
```

After doing t-test for each group, we can ALMOST reject H0, that is, dose will affect tooth growth. BUT, when comparing difference of tooth growth between 1 dose and 2 doses in OJ group, we failed to reject H0. So we might say, if we feed pigs orange juice, 1 or 2 doses may have little effect on tooth growth.

But, we may commit type I error when we seperate F-test into many t-tests. It will be better if we use ANOVA or regression to do it.