

Tooth Growth analysis

Xingmin Aaron Zhang

In this project, we will investigate the ToothGrowth dataset in R. The dataset was conducted in 60 guinea pigs. Each animal received one of three dose levels of vitamin C by one of two delivery methods, orange juice (OJ) or ascorbic acid (VC). The length of odontoblasts are measured as an indicator of tooth growth.

Load Data and Exploratory analysis

First, we load this dataset to our working directory.

```
library(datasets)
ToothGrowth$dose = as.factor(ToothGrowth$dose)
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 ...
## $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 ...
```

There are two values in supp column (OJ and VC) and three dose levels (0.5, 1, 2). Their combinations produced 6 experimental groups, with each group having 10 subjects.

```
library(dplyr)
ToothGrowth %>% group_by(supp, dose) %>% summarize(n = n())

## # A tibble: 6 x 3
## # Groups:   supp [?]
##   supp dose     n
##   <fct> <fct> <int>
## 1 OJ    0.5     10
## 2 OJ    1       10
## 3 OJ    2       10
## 4 VC    0.5     10
## 5 VC    1       10
## 6 VC    2       10
```

We use histograms to show how does length changes with supplement, dose, or the combination.

```
library(ggplot2)
library(gridExtra)
p <- ggplot(ToothGrowth)
p1 <- p + geom_boxplot(aes(x = supp, y = len))
p2 <- p + geom_boxplot(aes(x = dose, y = len))
p3 <- p + geom_boxplot(aes(x = interaction(dose, supp), y = len)) +
  xlab("experimental group")
grid.arrange(p1, p2, p3, ncol = 2, layout_matrix = rbind(c(1, 2), c(3, 3)))
```

From the above analysis, we can summarize our findings as follows: 1. There are two supplements, OJ and VC, and three doses, 0.5, 1 and 2. 2. There are 6 combinations of treatments, each supplement was used at three doses. Each combination had 10 replicates. 3. For different supplements, odontoblast length with OJ seems higher than VC. 4. odontoblasts length increased with increasing doses.

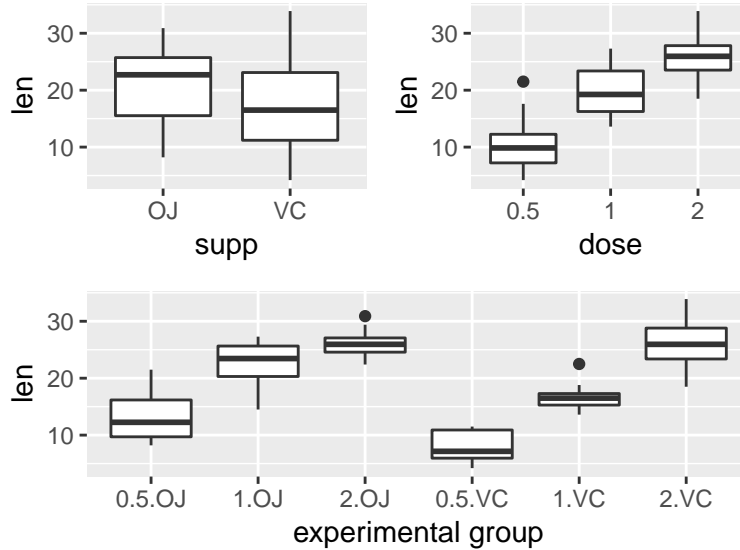


Figure 1: Figure 1 The length of tooth growth in response to different supplements at increasing doses

Statistical analysis

We want to perform Student's T test to test whether odontoblasts length was significantly different between each experimental group. In particular, we hope to test whether odontoblasts length is significantly higher when taken orange juice than ascorbic acid at each dose level, and whether growth is significantly higher at higher doses.

```

ToothGrowth$supp <- relevel(ToothGrowth$supp, "VC")
s <- split(ToothGrowth$len, interaction(ToothGrowth$supp, ToothGrowth$dose))

t.test.p <- function(x, ...) {
  result = t.test(x, ...)
  result$p.value
}

p_matrix <- sapply(s, function(x) sapply(s, function(y) t.test.p(y, x, alternative = "less")))
p_matrix[lower.tri(p_matrix, diag = TRUE)] <- NA
print(formatC(p_matrix, format = "e", digits = 2), quote = FALSE)

```

```

##      VC.0.5 OJ.0.5  VC.1    OJ.1    VC.2    OJ.2
## VC.0.5  NA    3.18e-03 3.41e-07 1.83e-08 2.34e-08 6.81e-12
## OJ.0.5  NA      NA    2.30e-02 4.39e-05 3.60e-06 6.62e-07
## VC.1    NA      NA      NA    5.19e-04 4.58e-05 1.18e-07
## OJ.1    NA      NA      NA      NA    4.83e-02 1.96e-02
## VC.2    NA      NA      NA      NA      NA    5.18e-01
## OJ.2    NA      NA      NA      NA      NA      NA

```

Because we are doing multiple t test, we hope to control false discovery rate by Bonferroni correction. We performed 15 t tests (after removing duplicates, i.e. t tests between a and b, or b and a, are considerate duplicates). Here we set false discovery rate 10%

```

alpha = 0.10
result_matrix <- ifelse(is.na(p_matrix), "-",
                        ifelse(p_matrix < alpha / 15,
                              formatC(p_matrix, format = "e", digit = 2), "NS"))
rownames(result_matrix) <- c("VC 0.5", "OJ 0.5", "VC 1", "OJ 1", "VC 2", "OJ 2")
colnames(result_matrix) <- c("VC 0.5", "OJ 0.5", "VC 1", "OJ 1", "VC 2", "OJ 2")
print(result_matrix, quote = FALSE)

```

```

##          VC 0.5 OJ 0.5   VC 1    OJ 1    VC 2    OJ 2
## VC 0.5 -      3.18e-03 3.41e-07 1.83e-08 2.34e-08 6.81e-12
## OJ 0.5 -      -        NS      4.39e-05 3.60e-06 6.62e-07
## VC 1   -      -        -        5.19e-04 4.58e-05 1.18e-07
## OJ 1   -      -        -        -        NS      NS
## VC 2   -      -        -        -        -        NS
## OJ 2   -      -        -        -        -        -

```

Conclusions

Assume that the samples are representative of each experimental group, follow normal distributions and that have same variations, we can conclude at 10% false discovery rate that:

1. at dose level 0.5, orange juice was significantly better at promoting tooth growth.
2. at dose level 1, orange juice was significantly better at promoting tooth growth.
3. at dose level 2, orange juice was no better than ascorbic acid in promoting tooth growth.
4. for ascorbic acid, increasing from dose 0.5 to 1, or from 1 to 2, increased tooth growth.
5. for orange juice, increasing from dose 0.5 to 1 increased tooth growth, but increasing from 1 to 2 did not.

All above considered, orange juice at dose 1 seems to be the best choice (lowest dose level with highest effect) in promoting tooth growth.