

Part2: Basic Inferential Data Analysis on ToothGrowth Dataset

Yunpeng Cheng

9/15/2020

Overview

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

1 Exploratory Data Analyses

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
library(datasets)
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 ...
```

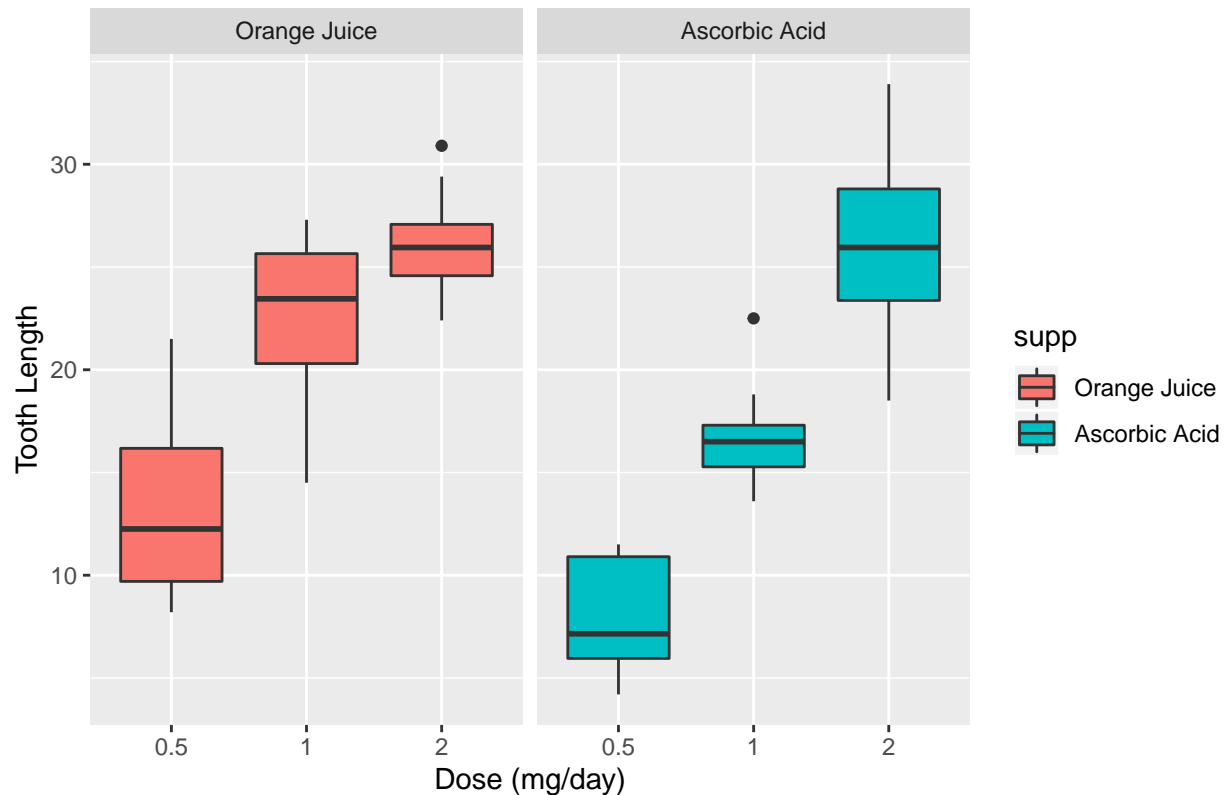
```
tg <- ToothGrowth
```

There are two kinds of supplements, the tooth length is corresponding to supplements and dose.

```
levels(tg$supp) <- c("Orange Juice", "Ascorbic Acid")
g <- ggplot(tg, aes(x=factor(dose), y=len))
g <- g+ facet_grid(.~supp)
g <- g+ geom_boxplot(aes(fill=supp))
g <- g+ labs(title="Tooth Length with Dosage for Each Supplement")
g <- g+ labs(x="Dose (mg/day)", y="Tooth Length")

print(g)
```

Tooth Length with Dosage for Each Supplement



From the boxplot, for 0.5 and 1 mg/day dosages, the average tooth length of orange juice is higher than that of ascorbic acid, which means the orange juice works better at those dosages. For 2 mg/day dosage, both averages are the same, but the variance of orange juice is lower.

2 Summary of ToothGrowth Dataset

Head of the dataset.

```
head(tg)
```

```
##      len      supp dose
## 1  4.2 Ascorbic Acid 0.5
## 2 11.5 Ascorbic Acid 0.5
## 3  7.3 Ascorbic Acid 0.5
## 4  5.8 Ascorbic Acid 0.5
## 5  6.4 Ascorbic Acid 0.5
## 6 10.0 Ascorbic Acid 0.5
```

Summary of the dataset.

```
summary(tg)
```

```
##      len      supp      dose
##  Min.   : 4.20   Orange Juice :30   Min.   :0.500
## 1st Qu.:13.07   Ascorbic Acid: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
```

3 Hypothesis Test

3.1 Hypothesis test on delivery method

Null hypothesis: for dose x mg/day, the delivery method have same effect on tooth growth, significance level is 5%.

- For dose x= 0.5 mg/day:

```
dose05 <- t.test(len ~ supp, data=subset(tg, dose==0.5))
dose05$conf.int
```

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

```
dose05$p.value
```

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

- For dose x= 1 mg/day:

```
dose1 <- t.test(len ~ supp, data=subset(tg, dose==1))
dose1$conf.int
```

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

```
dose1$p.value
```

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

- For dose x= 2 mg/day:

```
dose2 <- t.test(len ~ supp, data=subset(tg, dose==2))
dose2$conf.int
```

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

```
dose2$p.value
```

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

From above P value and confidence interval, for dose=0.5 and 1 mg/day, the null hypothesis can't be rejected, which means the delivery method have same effect on tooth growth.

But for dose=2 mg/day, the null hypothesis should be rejected. which means the delivery method have different effect on tooth growth.

These conclusions can be verified from the box plot.

3.2 Hypothesis test on dose

Null hypothesis: for same delivery method, different dose have same effect on tooth growth, significance level is 5%. Taking the data with orange juice method as an example.

Preparing the data.

```

supp_oj <- subset(tg, supp=="Orange Juice")
summary(supp_oj)

```

```

##           len           supp           dose
##  Min.      : 8.20   Orange Juice :30   Min.      :0.500
##  1st Qu.:15.53   Ascorbic Acid: 0   1st Qu.:0.500
##  Median :22.70                                Median :1.000
##  Mean      :20.66                                Mean      :1.167
##  3rd Qu.:25.73                                3rd Qu.:2.000
##  Max.      :30.90                                Max.      :2.000

```

```

supp_oj_051 <- subset(supp_oj, dose %in% c(0.5, 1.0))
supp_oj_052 <- subset(supp_oj, dose %in% c(0.5, 2.0))
supp_oj_12 <- subset(supp_oj, dose %in% c(1.0, 2.0))

```

```

summary(supp_oj_051)

```

```

##           len           supp           dose
##  Min.      : 8.20   Orange Juice :20   Min.      :0.50
##  1st Qu.:13.38   Ascorbic Acid: 0   1st Qu.:0.50
##  Median :18.65                                Median :0.75
##  Mean      :17.96                                Mean      :0.75
##  3rd Qu.:23.38                                3rd Qu.:1.00
##  Max.      :27.30                                Max.      :1.00

```

- For dose x= 0.5 and 1 mg/day:

```

supp_oj_051_test <- t.test(len ~ dose, paired=F, var.equal=F, data=supp_oj_051)
supp_oj_051_test$conf.int

```

```

## [1] -13.415634 -5.524366
## attr("conf.level")
## [1] 0.95

```

```

supp_oj_051_test$p.value

```

```

## [1] 8.784919e-05

```

- For dose x= 0.5 and 2 mg/day:

```

supp_oj_052_test <- t.test(len ~ dose, paired=F, var.equal=F, data=supp_oj_052)
supp_oj_052_test$conf.int

```

```

## [1] -16.335241 -9.324759
## attr("conf.level")
## [1] 0.95

```

```

supp_oj_052_test$p.value

```

```

## [1] 1.323784e-06

```

- For dose x= 1 and 2 mg/day:

```

supp_oj_12_test <- t.test(len ~ dose, paired=F, var.equal=F, data=supp_oj_12)
supp_oj_12_test$conf.int

```

```

## [1] -6.5314425 -0.1885575
## attr("conf.level")
## [1] 0.95

```

```
supp_oj_12_test$p.value
```

```
## [1] 0.03919514
```

All of the confidence intervals exclude 0, and p value is always less than 0.5, so the null hypothesis should be rejected. Hence different dose have different effect on the tooth growth.

4 Conclusions and Assumptions

Conclusions:

- The delivery method don't have obvious different effect on tooth growth.
- The dose have significant influence on tooth growth.
- 2 mp/day dose has larger effect on tooth growth than 0.5 and 1 mp/day dose.

Assumptions:

- The samples are not paired.
- Variance in each group are different.
- All samples are randomly selected.