The Course Project for Statistical Inference_Part 2

Pengfei LI

7/5/2020

Overview

This report aims to investigate a dataset involving the influence factor for tooth growth through several data analysis procedures. The data used in the following section is offered by R package "datasets", recording observations of the effects of VC on tooth growth among tested Guinea pigs. The dataset contains the data of the length of odontoblasts of 10 Guinea pigs in each tested group.

The experiments are separated into several parts through the dose level of Vitamin C, which is 0.5, 1 and 2 mg. With two kinds of delivery method, orange juice and ascorbic acid, the results are kept in the datasets, which will be analyzed in the following sections.

Data Pre-processing

Starting from the data pre-processing procedure, the code for the preparation of data analysis is offered as follows.

```
##Loading the data
library(datasets)
data("ToothGrowth")

##Pre-checking the data
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 ...
```

```
dim(ToothGrowth)
```

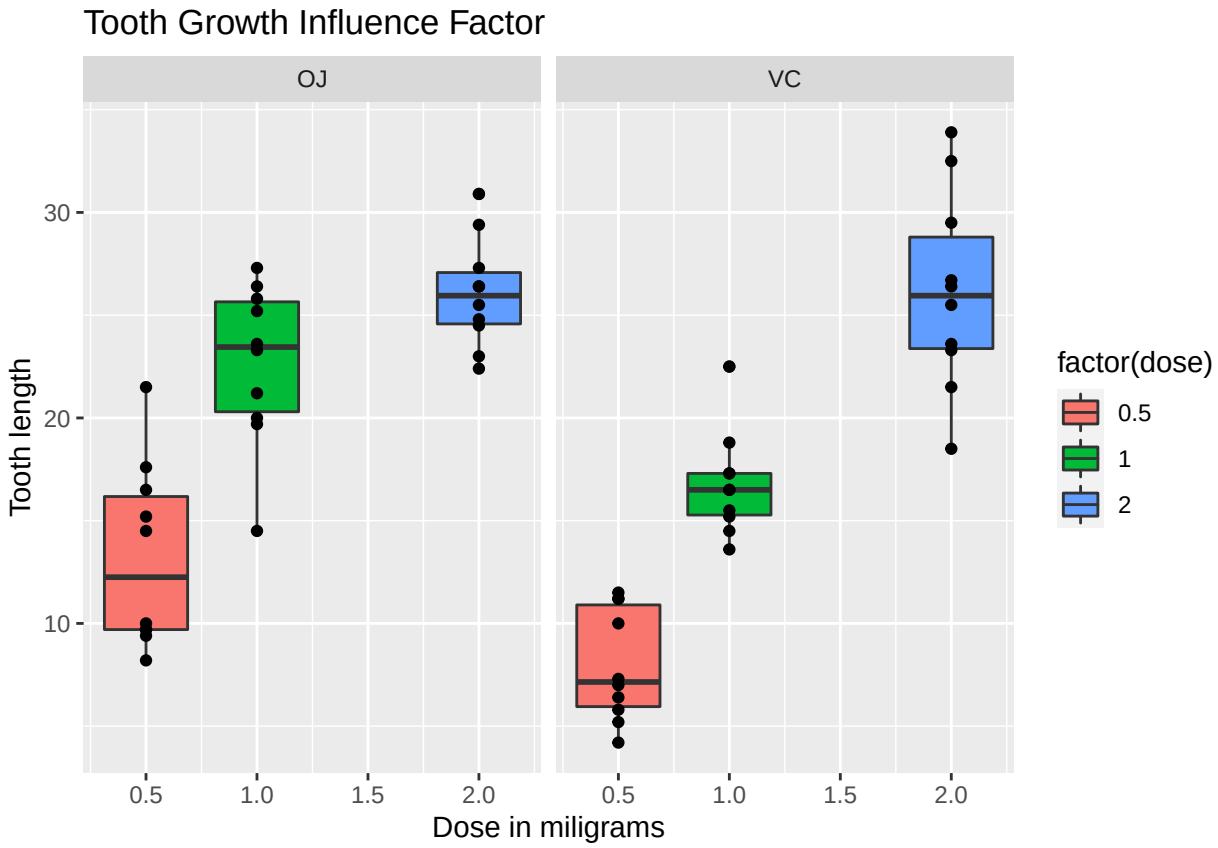
```
## [1] 60  3
```

```
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
```

```
library(ggplot2)
boxplot_overview <- ggplot(ToothGrowth, aes(x=dose, y=len)) +
        geom_boxplot(aes(fill=factor(dose)))+ geom_point()+
        facet_grid(.~supp)+labs(title = "Tooth Growth Influence Factor",x="Dose in miligrams",y="Tooth
boxplot_overview
```



The summary for the dataset of tooth growth shows that 30 observations are associated with the orange juice (AKA."OJ"), the rest are about ascorbic acid (AKA."VC"). The mean, median and maximum values are shown in the summary statistics.

Hypothesis 1: The delivery methods for Tested liquid

Based on the summary statistics in previous section, here comes with the first research question relevant with datasets, that is, whether the delivery methods of two experiment group have impacts on the tooth growth of Guinea pigs.

Assumptions

I assume that the delivery methods for two different tested liquids should have impacts on the experiment results.

Hypothesis

For the further analysis, I ease the question into two different hypothesis.

H0: The delivery method has some impacts on the tooth growth. H1: The delivery method has no impacts on the tooth growth.

Analysis

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"],
       ToothGrowth$len[ToothGrowth$supp == "VC"],
       paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

Results

From the upper results, we could see that the p-value is 0.06, while the confidence interval contains zero. In this case, p-value is higher than 0.05 (significant level), as well as the confidence interval contains zero. It means that the delivery methods have no influences on the tooth growth.

Hypothesis 2: The dose level for each tested group

As the upper analysis exclude the impacts of the delivery methods, the improved research issue should be about the impacts of high and low dose level for each tested group. In the experiment, the dose levels contain three layers, that is, 0.5mg, 1mg and 2mg. In this analysis section, we run analysis on 0.5mg and 2mg.

Assumption

In this section, I assume that the dose levels of each tested group are associated with the experiment results.

Hypothesis

Here I ease the question to two hypotheses as follows.

H0: Higher dose level(2mg, 1mg) has more impacts on the tooth growth than lower level(0.5mg).

H1: Higher dose level(2mg, 1mg) has less impacts on the tooth growth than lower level(0.5mg).

Analysis

```
t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 0.5],
       paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83648 18.15352
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```
t.test(ToothGrowth$len[ToothGrowth$dose == 1],
       ToothGrowth$len[ToothGrowth$dose == 0.5],
       paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Results

The upper analysis shows that p-values from two T-test results are both lower than 0.5, which both of the confidence intervals contain no zero. It means that null hypothesis is rejected and the higher dose level do have positive impacts on tooth growth of Guinea pigs.

Conclusion

In previous sections, the data analysis shows that the delivery methods of this experiment has no impacts while the high dose level of Vitamin C intake volume do affect the tooth growth of Guinea pigs.