Statistical Inference Course Project - Part II

Pingping

7/12/2020

Introduction

In this project, we first load the ToothGrowth data from the R datasets package, and did some basic data analyses. We also used confidential intervals and hypothesis tests to compare the tooth growth by supp and dose.

ToothGrowth data

```
toothdata <- ToothGrowth # get the ToothGrowth data into a data.frame
str(toothdata) # a brief summary of the toothdata
```

```
## '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 ...
```

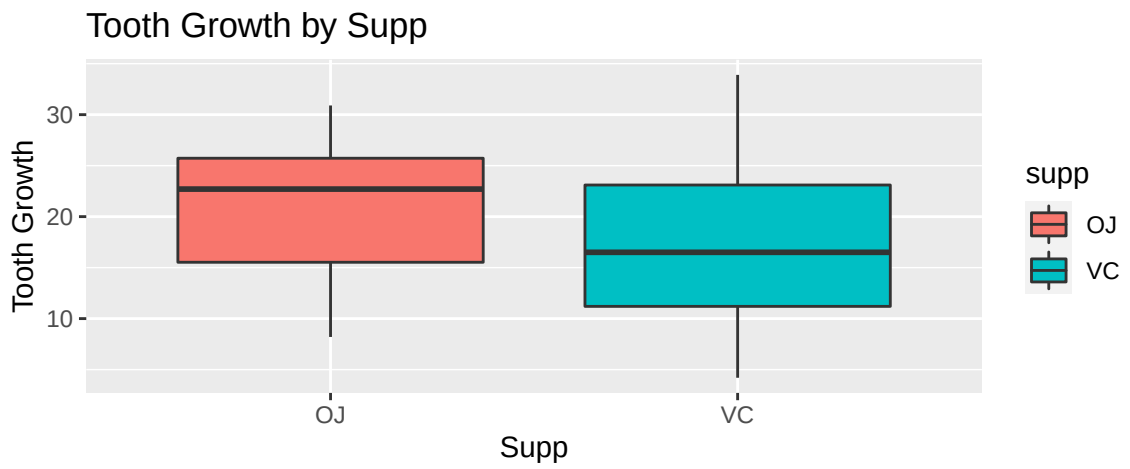
```
sum(!complete.cases(toothdata))
```

```
## [1] 0
```

Exploratory data Analyses

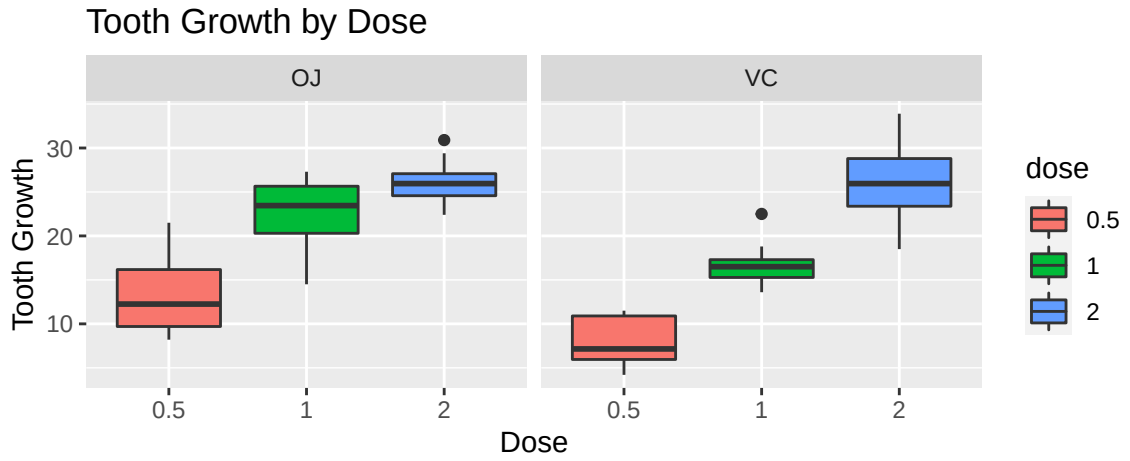
The following figure showed the tooth growth by different supplyments, namely VC and OJ.

```
g2 <- ggplot(toothdata, aes(x=supp, y=len,fill = supp)) +  geom_boxplot()
g2 <- g2 + labs(title = "Tooth Growth by Supp", x = "Supp", y = "Tooth Growth")
g2
```



The following figure showed the tooth growth by different dosages.

```
toothdata$dose <- as.factor(toothdata$dose)
g1 <- ggplot(toothdata, aes(x=dose, y=len,fill=dose)) +  geom_boxplot() + facet_grid(.~ supp)
g1 <- g1 + labs(title = "Tooth Growth by Dose", x = "Dose", y = "Tooth Growth")
g1
```



Summary The ToothGrowth data has 60 observations, decribing the growth of tooth using 2 kings of supplyments, VC and OJ, with three different dosages. Overall, OJ is more effient in assisting tooth growth. For both VC and OJ, the tooth showed a larger growth with a larger supplyment dosage.

Hypothesis test

Compare tooth growth by supp H_01: VC and OJ have similar influence on tooth growth

```
VC_data <- subset(toothdata, supp == "VC") # get the VC data
OJ_data <- subset(toothdata, supp == "OJ") # get the OJ data
t.test(VC_data$len, OJ_data$len, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  VC_data$len and OJ_data$len
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Summary: The p_value of the t.test is **0.06063**, which is bigger than 0.05. Therefore, we can not reject the null hypothesis, which means VC and OJ have similar influence on tooth growth. Moreover, the 95% confidence interval is frim -7.57 to 0.17, including 0, which also means the two supplyments have similar effects on tooth growth.

Compare supp dose H_02: Different doses have similar influence on tooth growth

(1) First, test the influence of different doses of VC.

```
VC0.5 <- subset(VC_data, dose == 0.5) # get the 0.5 dose data
VC1 <- subset(VC_data, dose == 1) # get the 1 dose data
VC2 <- subset(VC_data, dose == 2) # get the 2 dose data
t.test(VC0.5$len, VC1$len, paired = FALSE)$p.value
```

```
## [1] 6.811018e-07
```

```
t.test(VC1$len, VC2$len, paired = FALSE)$p.value
```

```
## [1] 9.155603e-05
```

As we can see, both of the p-value are far less than 0.05, suggesting that we are able to reject the Null hypotheses.Therefore, there are differences in the influence of different doses of VC in tooth growth.

(2) Then, test the influence of different doses of OJ

```
OJ0.5 <- subset(OJ_data, dose == 0.5) # get the 0.5 dose data
OJ1 <- subset(OJ_data, dose == 1) # get the 1 dose data
OJ2 <- subset(OJ_data, dose == 2) # get the 2 dose data
t.test(OJ0.5$len, OJ1$len, paired = FALSE)$p.value
```

```
## [1] 8.784919e-05
```

```
t.test(OJ1$len, OJ2$len, paired = FALSE)$p.value
```

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

As we can see, both of the p-value are far less than 0.05, suggesting that we are able to reject the Null hypotheses.Therefore, there are differences in the influence of different doses of OJ in tooth growth.

Summary: VC and OJ have similar influence on tooth growth. For both VC and OJ, the larger the dose is, the more efficient in assisting tooth growth.