Statistical Inference Course Final Project - Part 2

André Marinho

05/08/2020

0. Overview

This is the second part of project report from Coursera Statistical Inference Course. The taks for this part are following:

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. Loading Dataset and Performing EDA

```
data_toothGrowth <- data("ToothGrowth")

# EDA: get dataset info
?ToothGrowth

# EDA: data exploration
library(ggplot2)

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
```

```
dim(ToothGrowth)
```

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

```
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 ...
```

```
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6 14.5
## [16] 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6  9.7  8.2
## [31]  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4 23.0
```

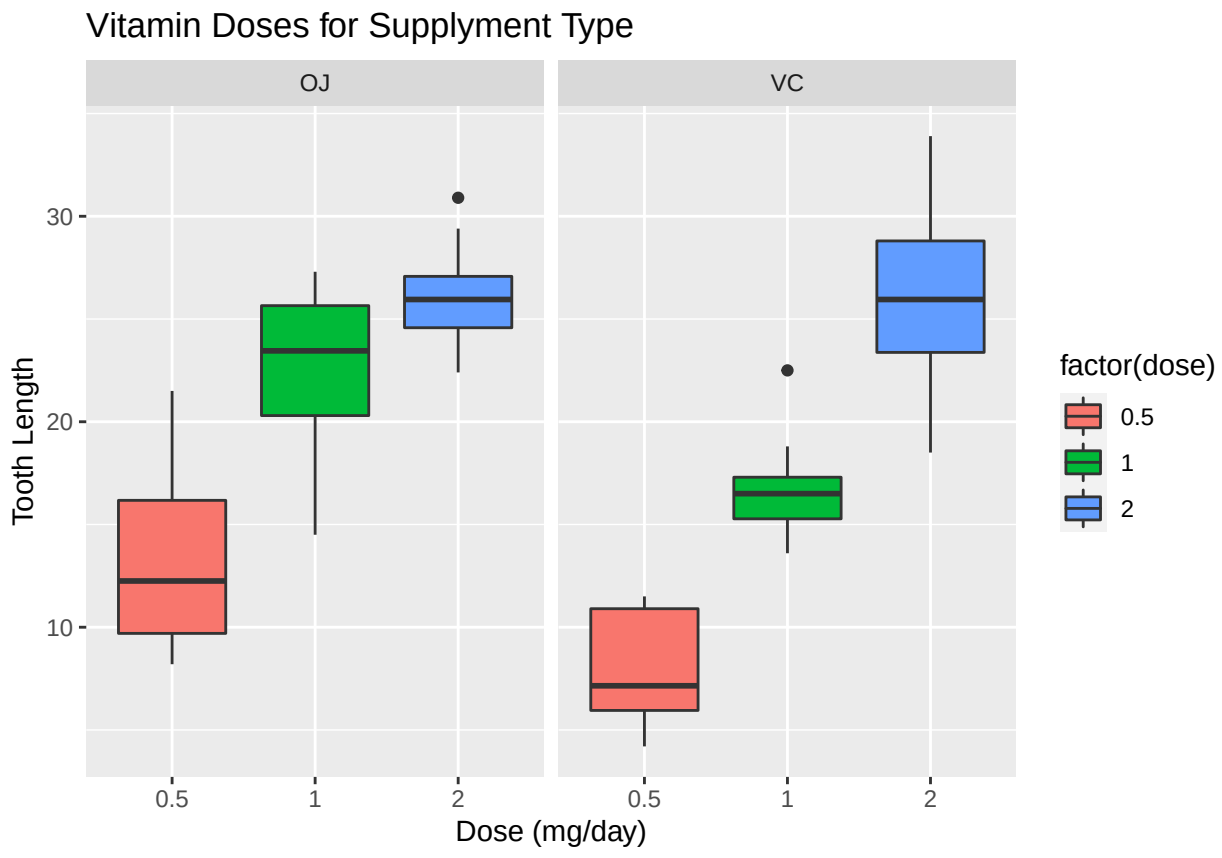
```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
plot_ToothGrowth <- ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
        geom_boxplot() + facet_grid(.~supp) +
        labs(title="Vitamin Doses for Supplyment Type", x="Dose (mg/day)", y="Tooth Length")
plot_ToothGrowth
```



2. Summary of the Data

```
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
```

3. Getting Confidence Interval

```
dose05 <- subset(ToothGrowth, dose==0.5)
dose1 <- subset(ToothGrowth, dose==1.0)
dose2 <- subset(ToothGrowth, dose==2.0)

test05 <- t.test(len ~ supp, data=dose05, paired=FALSE, var.equal=FALSE)
test1 <- t.test(len ~ supp, data=dose1, paired=FALSE, var.equal=FALSE)
test2 <- t.test(len ~ supp, data=dose2, paired=FALSE, var.equal=FALSE)

intervals_pValue <- cbind(test05$p.value, test1$p.value, test2$p.value)

intervals_lower_bound <- cbind(test05$conf.int[1], test1$conf.int[1], test2$conf.int[1])

intervals_upper_bound <- cbind(test05$conf.int[2], test1$conf.int[2], test2$conf.int[2])

confidence_interval_results <- data.frame(
        "p.value" = c(intervals_pValue),
        "Conf.Low" = c(intervals_lower_bound),
        "Conf.High" = c(intervals_upper_bound),
        row.names = c("Dose 0.5 ->", "Dose 1.0 ->", "Dose 2.0 ->")
)
confidence_interval_results
```

```
##                  p.value  Conf.Low Conf.High
## Dose 0.5 -> 0.006358607  1.719057  8.780943
## Dose 1.0 -> 0.001038376  2.802148  9.057852
## Dose 2.0 -> 0.963851589 -3.798070  3.638070
```

4. Assumptions and Conclusions

Assumptions:

1. Tidy dataset.
2. Tooth growth differ based on supplement type and dosage.

Conclusions:

1. Tooth growth data follows a normal distribution.
2. For Doses 0.5mg/day and 1.0mg/day, it rejects the null hypothesis (p-value is less than 5%), meaning Orange Juice (OJ) and Ascorbic Acid (VC) differ significantly at lower doses.
3. For Dose 2.0mg/day, it fails to reject the null hypothesis (p-value is greather than 5%), meaning OJ and VC does not differ significantly at higher doses.
4. Increasing the dose of JC or VC increases the tooth growth in guinea pigs.