

Statistical Inference Course Project Part 2

jacethedatascientist

August 12, 2019

Statistical Inference Course Project Part 2

Jace Galleon

I. Overview

This is the second part of the Statistical Inference Course Project from Coursera.

In this project, we're going to analyze the ToothGrowth data in the R datasets package.

II. Objectives

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.
-

III. Data Processing

A. Exploration

```
data("ToothGrowth")
print(names(ToothGrowth))
```

```
## [1] "len" "supp" "dose"
```

```
print(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
```

```
print(paste("Number of Rows: ", nrow(ToothGrowth)))
```

```
## [1] "Number of Rows: 60"
```

```
print(sapply(ToothGrowth, class))
```

```
##      len      supp      dose  
## "numeric" "factor" "numeric"
```

B. Preprocessing

1. Since the dosage is a level not an actual value, it will be converted to factor.

```
procData <- ToothGrowth  
procData$dose <- as.factor(procData$dose)
```

2. Relationship between the Tooth Length and Supplements

```
##      OJ      VC  
## 20.66333 16.96333
```

3. Relationship between the Tooth Length and Dosage Level

```
##      0.5      1      2  
## 10.605 19.735 26.100
```

4. Variance by Dosage Level

```
sapply(meanDose, var)
```

```
##      0.5      1      2  
## 20.24787 19.49608 14.24421
```

5. Variance by Supplements

```
sapply(meanSupp, var)
```

```
##      OJ      VC  
## 43.63344 68.32723
```

IV. Analysis

Tooth Growth Comparison

A. By Supplements

Hypotheses Testing

Null : *The supplements has no impact on Tooth Growth*

Alternative : *At least one of the supplements has impact on Tooth Growth*

The P-value is **0.06063** and is **greater than the 0.05 significance value**, then, *we don't have any sufficient data to reject the null hypothesis.*

Thus, both the supplements **has no impact** on Tooth Growth

B. By Dosage Level

Hypotheses Testing

Null : *Higher dosage level has no impact on Tooth Growth*

Alternative : *Higher dosage level has impact on Tooth Growth*

The P-value is **almost 0** and obviously **lesser than the 0.05 significance value**, then, *we have sufficient data to reject the null hypothesis.*

Thus, higher dosage level **has impact** on Tooth Growth.

V. Conclusion

Summary

By using the T-Test, we can now say that:

1. The supplements (Orange Juice and Vitamin C) **has no impact** on the Tooth Growth.
2. Higher dosage level (0.2 or higher) **has impact** on Tooth Growth

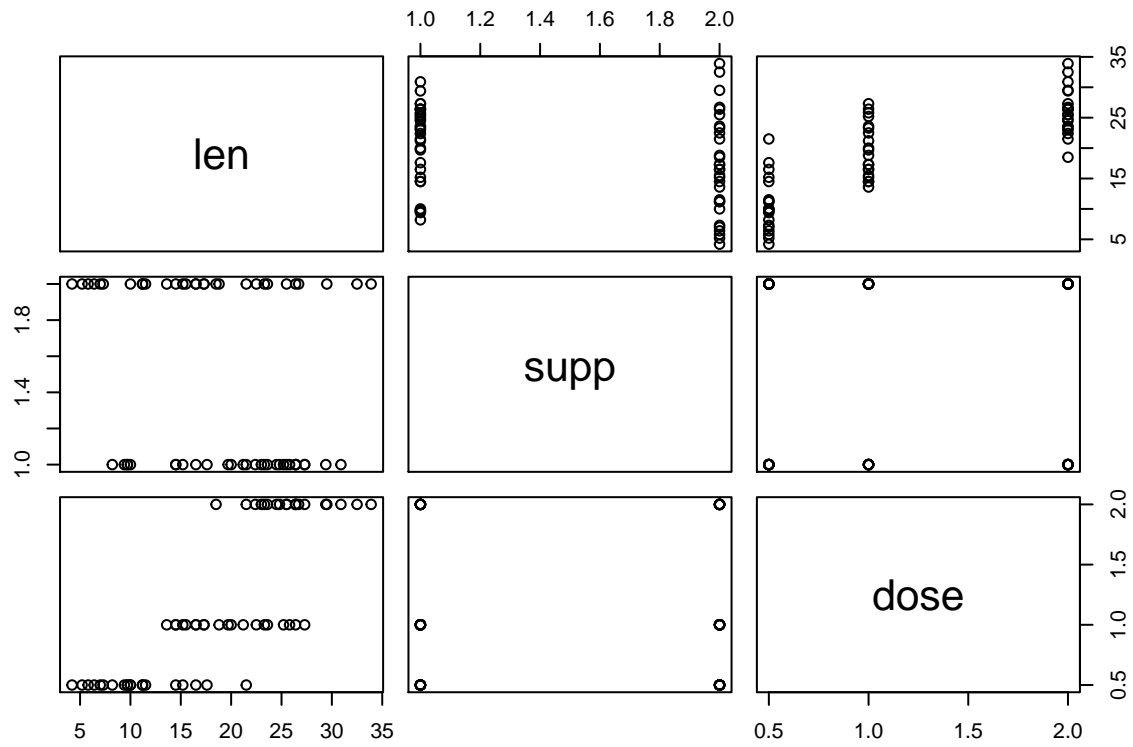
Assumptions

- each subject is randomly assigned
- each subject is a representative of the population
- observations are independent

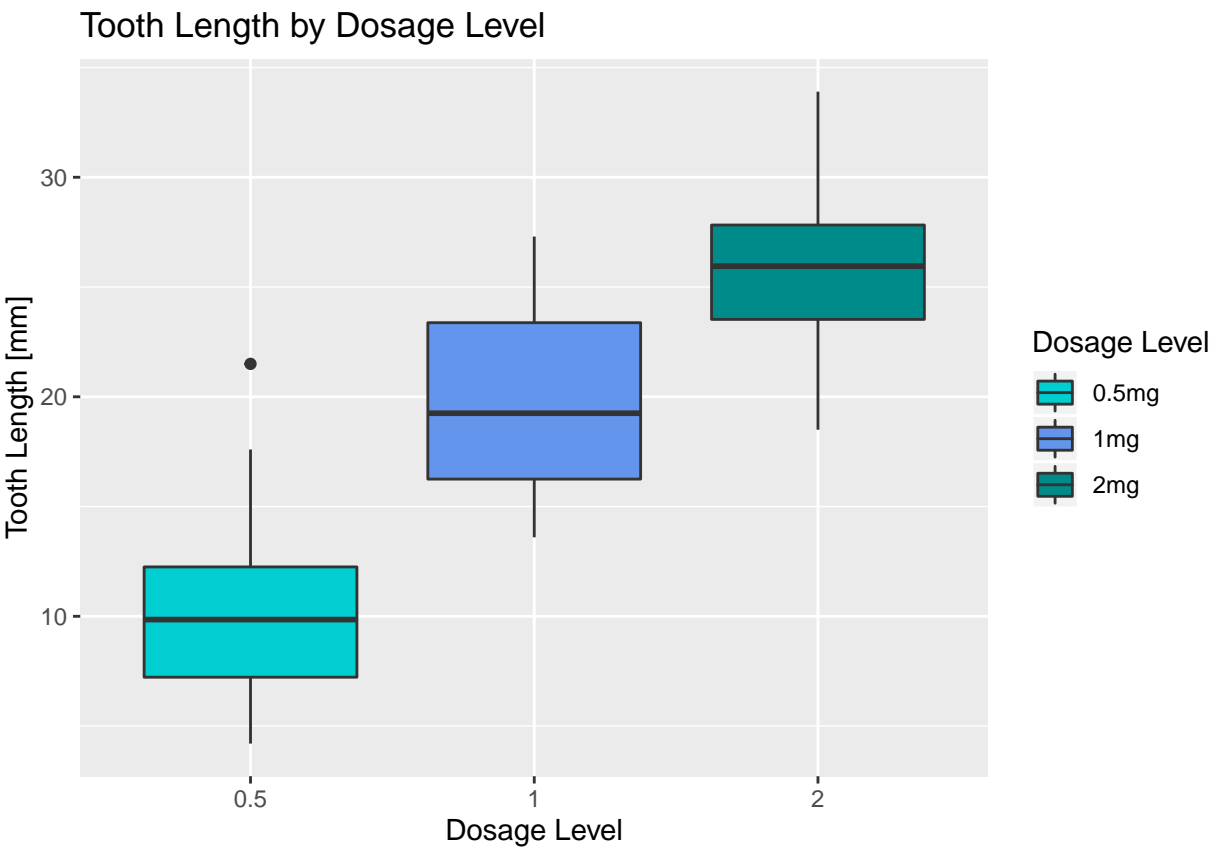
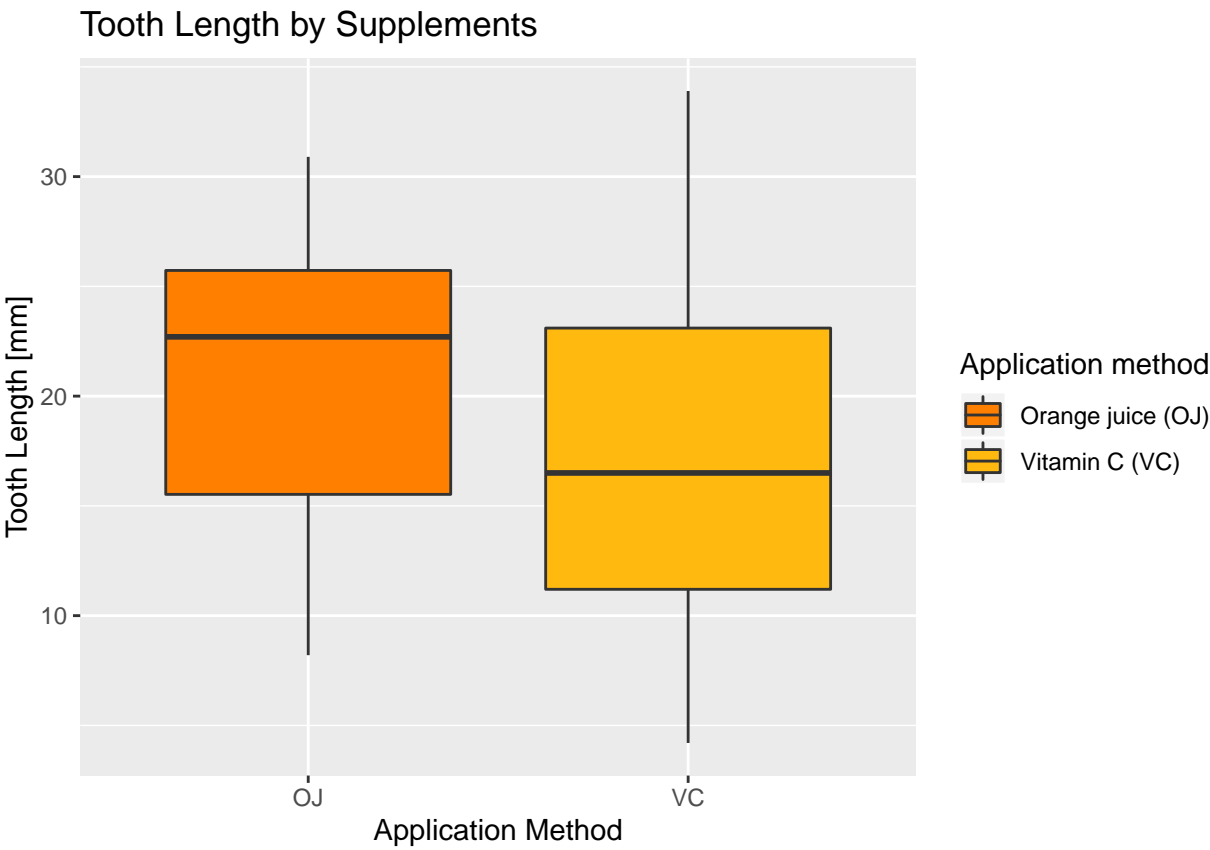
This formally ends the Part 2 of the Course Project. Thank You!

VI. Additional Chart

ToothGrowth Scatter Plot



ToothGrowth Box Plot



T-Test By Supplements

```
##
## Welch Two Sample t-test
##
## data:  procData$len[procData$supp == "OJ"] and procData$len[procData$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
```

T-Test By Dosage Level

```
##
## Two Sample t-test
##
## data:  procData$len[procData$dose == 2] and procData$len[procData$dose != 2]
## t = 7.0489, df = 58, p-value = 2.426e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 7.82616 14.03384
## sample estimates:
## mean of x mean of y
## 26.10 15.17
```