Statistical Inference Project 2 - Exploring Tooth Growth

Sukdev Koner

Sept 24, 2018

Overview

In this project, I am going to explore `TootGrowth` data of the R 'datasets' package by comparing the guinea pigs' tooth growth by supplement and dose. First, I would do some exploratory analysis on the data. Then I will use Hypothesis test to make conclusions about what influences the tooth growth.

The `ToothGrowth` data set consists of 60 observations and 3 variables:
1. `len`: Teeth length in millimeters (numeric variable)
2. `supp`: Supplement type (factor variable with levels VC and OJ)
3. `dose`: Dose in milligrams (numeric variable) [0.5, 1, and 2 mg]

Load ToothGrowth data and perform basic exploratory analysis

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

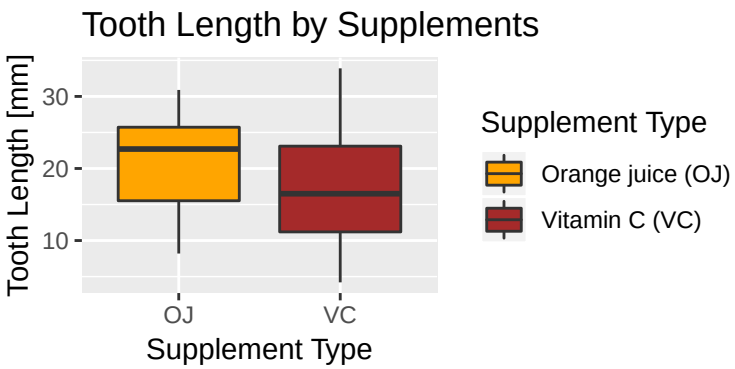
```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Here `dose` is a numeric variable but we have to convert to as **Factor** for subsequent inference analysis. Here we can see that for 'Supplement Type', it has 2 different levels and 'Dose Level' has three diferent levels.

A. Analysis of Tooth Length by Supplement Type :

```
ggplot(aes(x = supp, y = len), data = toothdata) +
geom_boxplot(aes(fill = supp)) + xlab("Supplement Type") +
ylab("Tooth Length [mm]") + ggtitle("Tooth Length by Supplements") +
scale_fill_manual(name = "Supplement Type", values = c("orange", "brown"), labels = c("Orange juice (OJ
```

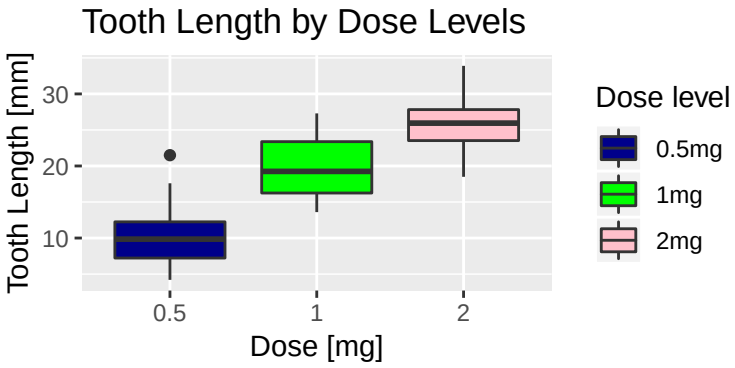


This box plot above shows that guinea

pigs which received orange juice (OJ) have a greater average tooth length than those pigs which got Vitamin C (VC).

B. Analysis of Tooth Length by Dose Level :

```
ggplot(aes(x = dose, y = len), data = toothdata) +
geom_boxplot(aes(fill = dose)) + xlab("Dose [mg]") +
ylab("Tooth Length [mm]") + ggtitle("Tooth Length by Dose Levels") +
scale_fill_manual(name = "Dose level", values = c("darkblue", "green", "pink"), labels = c("0.5mg", "1mg
```



This plot states that when we increase the Dose, we are seeing increase in Average(mean) tooth length. And the 'Range (Max - Min)' for Tooth Length for various 'Doses' are so spread out that they almost did not overlap.

Hypothesis tests to compare tooth growth by Supplement and Dose

1. By Supplement Type:

We will check if the Difference in 'Tooth length means' between pigs who received Orange Juice and who received Vitamin C, is statistically different from 0.

Hypothesis

Supplement type has no impact on tooth growth.
H_0: Both groups (OJ and VC) have the same mean.
H_A: Means are different.

We will use t-test with unequal variances of two samples.

```
t <- t.test(len ~ supp, data = toothdata, var.equal = FALSE)
t$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
t$p.value
```

```
## [1] 0.06063451
```

The p-value is greater than Significance Level($\alpha = 0.05$) and "Confidence Interval" of the test contains zero,so we **fail to reject Null Hypothesis**. Hence we can say that **'Supplement Types' seems to have no impact on Tooth growth.**

2. By Dose Level:

We will check if the Difference in 'Tooth length means' between pigs who received different Dose is statistically different from 0. We will first compare between **dose 2mg and dose 0.5mg**. After that, we will see the comparison test between **dose 1mg and dose 0.5mg** and then between **dose 2mg and dose 1mg**.

Hypothesis

Higher doses of supplement causes less tooth growth.
H_0: Mean of Dose level 2 is smaller or equal than dose level 0.5.
H_A: Mean of level 2 is greater than level 0.5.

```
# run t-test using dose amounts 0.5 and 2.0
t1 <- t.test(len~dose,data= subset(toothdata, toothdata$dose %in% c(0.5,2.0)))
Dose_half_Vs_2_Mg <- c(round(t1$conf.int,3), round(t1$p.value,15))
# run t-test using dose amounts 1.0 and 0.5
t2 <- t.test(len~dose,data= subset(toothdata, toothdata$dose %in% c(1.0,0.5)))
Test_type <- c("Conf. Interval-Low", "Conf. Interval-High","P-Value")
Dose_1_Vs_half_Mg <- c(round(t2$conf.int,3), round(t2$p.value,9))
# run t-test using dose amounts 2.0 and 1.0
t3 <- t.test(len~dose,data= subset(toothdata, toothdata$dose %in% c(2.0,1.0)))
Dose_2_Vs_1_Mg <- c(round(t3$conf.int,3), round(t3$p.value,9))
r1 <- rbind(Test_type,Dose_half_Vs_2_Mg,Dose_1_Vs_half_Mg,Dose_2_Vs_1_Mg)
r1
```

```
##                   [,1]                 [,2]                  [,3]
## Test_type         "Conf. Interval-Low" "Conf. Interval-High" "P-Value"
## Dose_half_Vs_2_Mg "-18.156"            "-12.834"             "4.4e-14"
## Dose_1_Vs_half_Mg "-11.984"            "-6.276"              "1.27e-07"
## Dose_2_Vs_1_Mg    "-8.996"             "-3.734"              "1.9064e-05"
```

After comparing all results, their p-values are very low (less than significance level) and the confidence intervals also do not contain zero, so we can **reject the null hypothesis** and conclude that **'Dose' does affect the tooth length**.

Conclusion

According the t-tests, we can conclude that
- **Supplement type** has no impact of thooth growth
- increasing **Dose Level** results in increasing tooth growth

Assumptions

Following assumptions are made here:
- each subject is randomly assigned and representative of population
- variables are independent and identically distributed (i.i.d.).

[Note : The complete assignment can be found on GitHub.]