SI_Project_P2

Rahul Jain

30/01/2022

Overview

The purpose of this analysis is to analyze Tooth Growth data-set by comparing tooth growth by supplement and dose. Exploratory data analysis on the given data set is the first step and then doing hypothesis tests to compare tooth growth by supp and dose

Loading the dataset and required libraries

```
library(datasets)
data("ToothGrowth")
library(ggplot2)
```

Exploratory Data Analysis

Dimnesions of the data-set There are 60 observations and 3 variables

```
dim(ToothGrowth)
```

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

Look at the structure of the data-set

```
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 ...
```

Look at first few rows of the data-set

```
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
```

Summary of the data-set

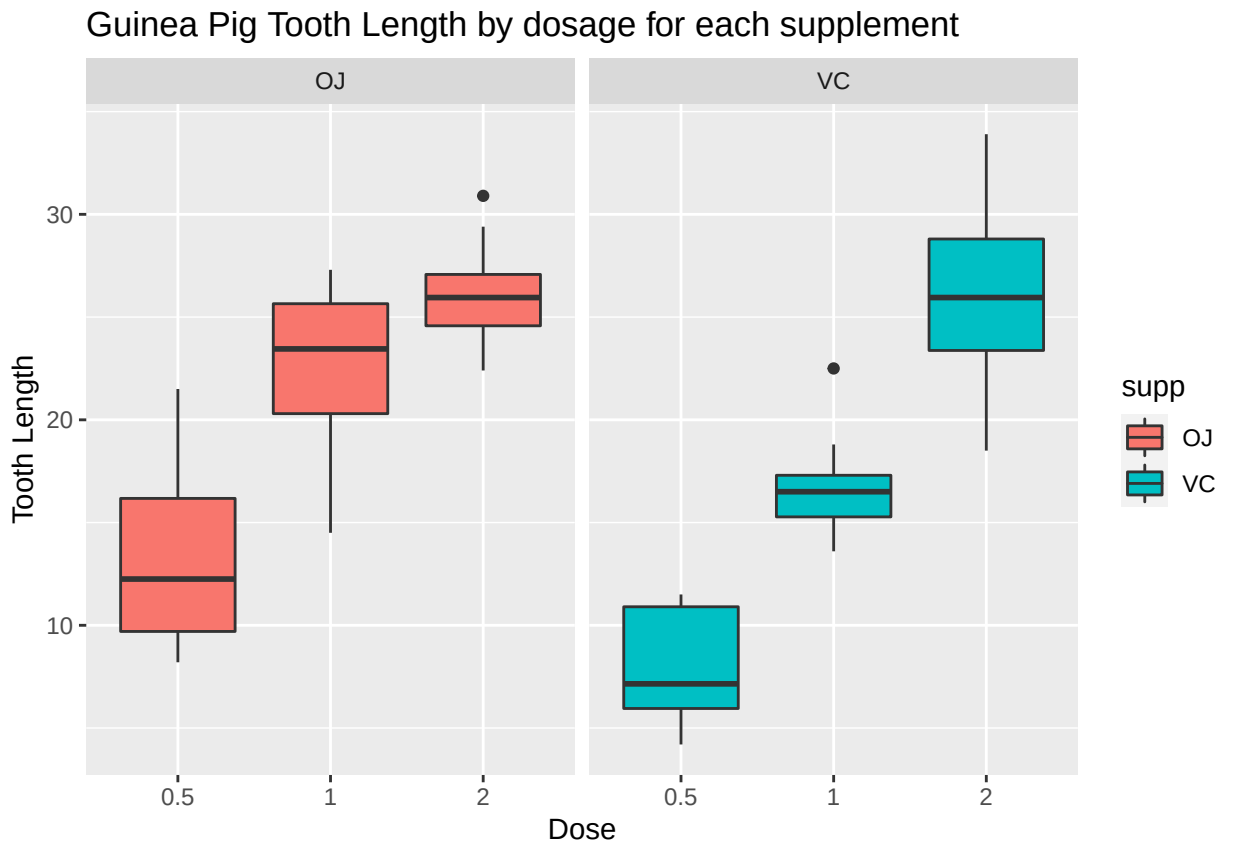
```
summary(ToothGrowth)
```

```
##       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
```

There are 30 observations for both supplement type

Boxplots to compare the growth with supplement type and dose The graph shows that increased dosage is effective in increasing tooth length whereas the results are similar for supplement type

```
g <- ggplot(ToothGrowth,aes(x=factor(dose),y=len))
g + facet_grid(.~supp) + geom_boxplot(aes(fill=supp)) +
    labs(x="Dose",y="Tooth Length",
         title = "Guinea Pig Tooth Length by dosage for each supplement")
```



Hypothesis Testing

```
t.test(len~supp,data = ToothGrowth)
```

1. Tooth Growth is similar for both supplement types

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Null - Hypothesis - Difference in tooth length for both supplements is 0

Alternate Hypothesis - Difference in tooth length for both supplements is not 0

From the t-test, I have below observations

- Mean difference between 2 groups = 3.7
- t-statistic is 1.9 which is not very high for 55 degrees of freedom
- p-value is more than 0.05 which means we cannot reject null hypothesis
- Confidence interval contains 0, hence, we cannot reject null

```
t.test(len~factor(dose),data = subset(ToothGrowth,dose %in% c(0.5,1)))
```

2. Increase in dosage from 0.5mg to 1mg results in increase in tooth length

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Null - Hypothesis - Difference in tooth length for both dosages(0.5 and 1) is 0

Alternate Hypothesis - Difference in tooth length for both dosages(0.5 and 1) is not 0

From the t-test, I have below observations

- Mean difference between 2 groups = 9.1
- t-statistic is -6.45 which is very high for 38 degrees of freedom, 97.5 probability requires t-statistic of around 2.02
- p-value is very less than 0.05 which means we can reject null hypothesis and accept alternate
- Confidence interval is very below 0, hence, we can reject null

```
t.test(len~factor(dose),data = subset(ToothGrowth,dose %in% c(1,2)))
```

3. Increase in dosage from 1 mg to 2 mg results in increase in tooth length

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(dose)
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Null - Hypothesis - Difference in tooth length for both dosages(1 and 2) is 0

Alternate Hypothesis - Difference in tooth length for both dosages (1 and 2) is not 0

From the t-test, I have below observations

- Mean difference between 2 groups = 6.4
- t-statistic is -4.9 which is very high for 37 degrees of freedom, 97.5 probability requires t-statistic of around 2.02
- p-value is very less than 0.05 which means we can reject null hypothesis and accept alternate
- Confidence interval is very below 0, hence, we can reject null

Conclusions and assumptions

Conclusions

- Increase in dosage leads to increase in tooth length
- Both supplements leads to similar increase in tooth length

Assumptions

- Tooth length follows normal distribution
- Other factors have been controlled while doing experiment