

Basic Inferential Data Analysis Instructions

vyndk

September 3, 2018

Overview

In this project, we shall analyze the ToothGrowth data in the R datasets package. # Load data and do some exploratory analysis Now we load the data and do some exploratory analysis.

```
library(tidyverse)
data("ToothGrowth")
# view size of data
dim(ToothGrowth)
```

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

```
# view some top data
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
```

```
# explore data type for each columns
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 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

We can see that dose may be a factor variable. So we convert it to factor

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

Basic summary of the data

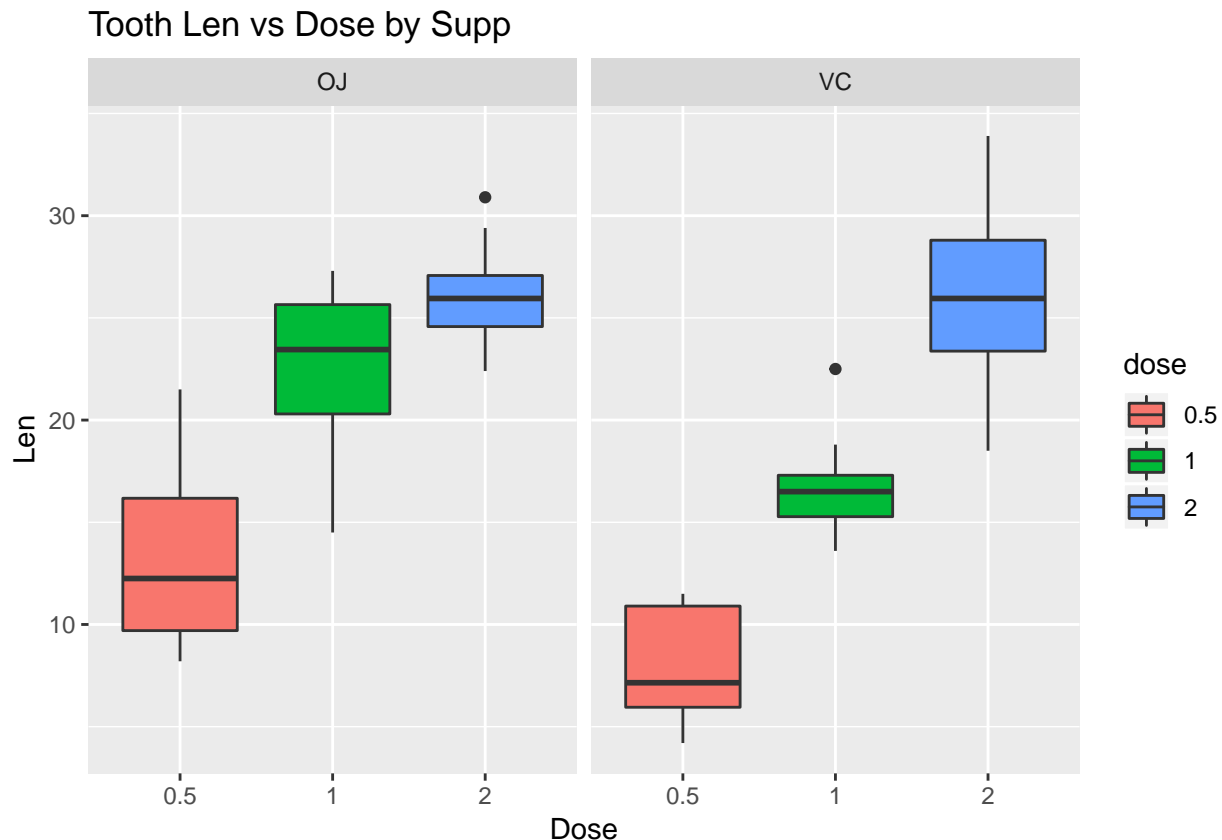
Now we summary the data

```
#summay data
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
```

We plot the tooth len by dose and group by supp

```
ggplot(aes(x=dose, y=len), data=ToothGrowth) +
  geom_boxplot(aes(fill=dose)) +
  labs(x = "Dose", y= "Len", title = "Tooth Len vs Dose by Supp") +
  facet_grid(. ~ supp)
```



Compare tooth growth by supp We shall compare tooth growth by supp We have 2 group: group 1 is tooth growth with OJ supp, group 2 is tooth growth with VC supp H_0 : two group have the same mean H : two group have different mean We apply T-test for this test

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

```
##
## 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 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
```

Here we can see that p-value greater than the significance level of 0.05 and the confidence interval contains zero. Therefore, it is not enough evidence to reject the H_0 . We can say that supp has no impact on tooth growth. ## Compare tooth growth by dose Compare tooth growth with each pair of dose: 0.5 and 1, 0.5 and 2, 1 and 2 1. Compare tooth growth with dose is 0.5 and 1

```
sub_data <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1))
t.test(len~dose,data=sub_data)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means 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
```

2. Compare tooth growth with dose is 0.5 and 2.0

```
sub_data <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2))
t.test(len~dose,data=sub_data)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -18.15617 -12.83383
## sample estimates:
## mean in group 0.5 mean in group 2
## 10.605 26.100
```

3. Compare tooth growth with dose is 1.0 and 2.0

```
sub_data <- subset(ToothGrowth, ToothGrowth$dose %in% c(1,2))
t.test(len~dose,data=sub_data)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means 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
```

We can see that all p-values are smaller than the significance level of 0.05 and the confidence interval of each test does not cross over zero. Therefore we can say that dose has effect on tooth grow

Conclusion

Given the assumptions that the sample is representative of the population and the observations are independent, we can conclude that supp has no effect on tooth growth while dose has impact on tooth growth.