

Statistical Inference Course Project

Part 2: Basic Inferential Data Analysis

In this second part of the project, the ToothGrowth data is analyzed.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(ggplot2)
library(datasets)
data(ToothGrowth)
#plot(x = ToothGrowth$supp, y=ToothGrowth$dose)

g <- ggplot(ToothGrowth, aes(x=len, y=dose, colour=supp)) +
  geom_smooth()+
  labs(x="Len", y=expression("Dose"))+
  labs(title = paste('Relation between length of teeth with dose'))+
  geom_line(aes(group=supp))
g
```



The above figure shows the relation between the growth of teeth(len) and dose of two different types(supp: VC, OJ).

2. Provide a basic summary of the data.

```
head(ToothGrowth,10)
```

```
##      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
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

```
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 ...
```

```
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. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

In the present test it is assumed that the groups are independent and different variance per group. From sixty observations, 30 corresponds to subjects treated with supp VC and another 30 with supp OJ at doses 0.5, 1 and 2. Let's see the influence of each treatment in the tooth length

```
# Comparing tooth growth by supp
g1 <- ToothGrowth$len[1 : 30] # supp : VC
g2 <- ToothGrowth$len[31 : 60] # supp : OJ
t.test(g2, g1, paired = FALSE, var.equal = TRUE, alternative = "two.sided")
```

```
##
## Two Sample t-test
##
## data:  g2 and g1
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
## 20.66333 16.96333
```

4. State your conclusions and the assumptions needed for your conclusions

From the previous results we can conclude that the higher values of tooth length are influenced by supp VC. The difference in means shows that the two groups under experiment have a different behaviour under the treatment with supp VC and OJ. Since the p-value = 0.06039 and $\alpha = 0.05$ for this two sided hypothesis test, we fail to reject the null hypothesis in this case.