

Statistical Inference Project, Part 2

Hoang Tam Vo

Overview

In this part of the project, I analyze the ToothGrowth data in the R datasets package. Specifically, I load the ToothGrowth data and perform some basic exploratory data analyses as well as provide a basic summary of the data. Further, confidence intervals and hypothesis tests are used to compare tooth growth by supp and dose. Finally, conclusions and the assumptions needed are discussed.

Questions 1: Load the ToothGrowth data and perform some basic exploratory data analyses

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

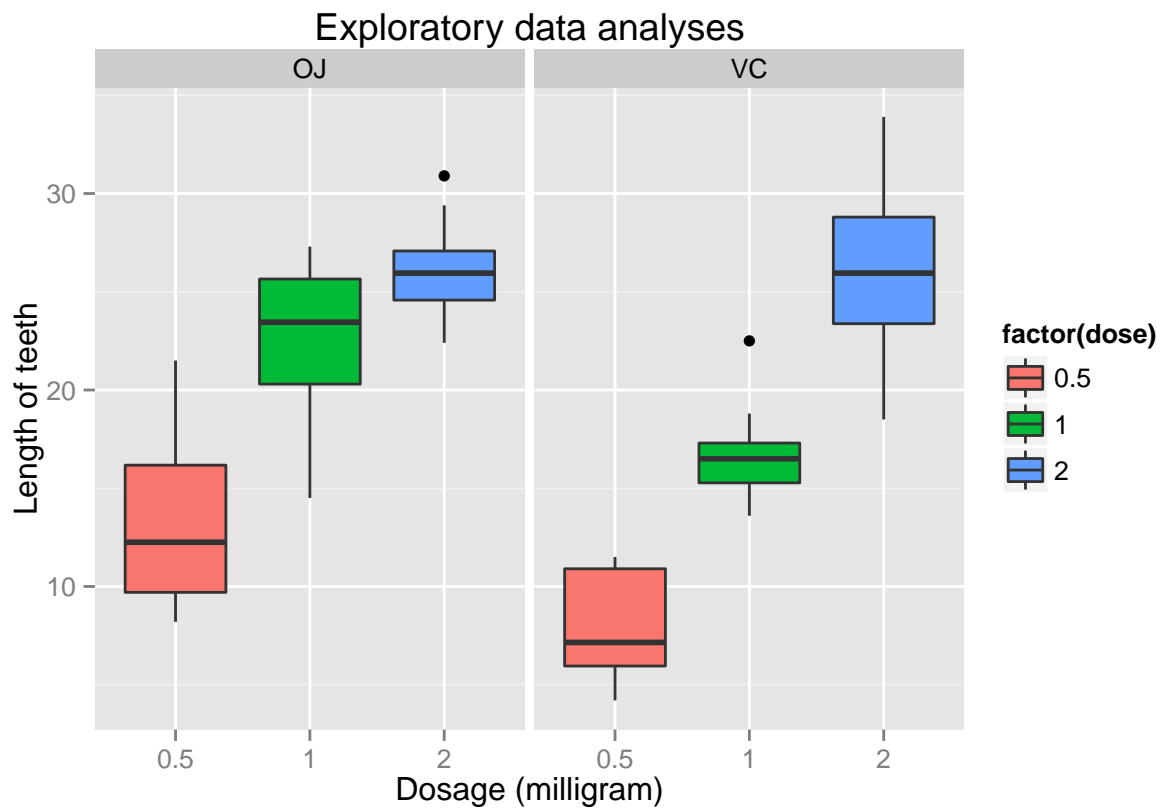
```
## 'data.frame':   60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
ToothGrowth[c(1:8),]
```

```
##      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
```

```
plot <- ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose)))

plot + geom_boxplot(notch=F) +
  facet_grid(.~supp) +
  scale_x_discrete("Dosage (milligram)") +
  scale_y_continuous("Length of teeth") +
  ggtitle("Exploratory data analyses")
```



Questions 2: Provide a basic summary of the data

```

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

```

```

##      len      supp  dose
##  Min.   : 4.2    OJ:30  0.5:20
##  1st Qu.:13.1    VC:30   1 :20
##  Median :19.2             2 :20
##  Mean   :18.8
##  3rd Qu.:25.3
##  Max.   :33.9

```

```

table(ToothGrowth$dose, ToothGrowth$supp)

```

```

##
##      OJ VC
##  0.5 10 10
##    1 10 10
##    2 10 10

```

Question 3: Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

```
test1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
test2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
result <- data.frame("p-value"=c(test1$p.value, test2$p.value),
                     "Low Confidence"=c(test1$conf[1],test2$conf[1]),
                     "High Confidence"=c(test1$conf[2],test2$conf[2]),
                     row.names=c("Equal","Unequal"))
result
```

```
##           p.value Low.Confidence High.Confidence
## Equal    0.06039      -0.167         7.567
## Unequal  0.06063      -0.171         7.571
```

Question 4: State conclusions and the assumptions needed

Given the above analysis, several conclusions can be drawn:

1. Orange juice (OJ) and vitamin C (VC) have different effect on the growth of the tooth.
2. There is a difference in the tooth growth when the doses are larger.

These analysis and conclusions are based on the following assumptions:

- The sample data are representative for the entire population.
- The distribution of the data is normal.