

Statistical Inference Course Project (Part 2)

Title: Analyze the ToothGrowth data

Overview

In this project we use the ToothGrowth data and provide a basic summary. We also use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose and describe a conclusion.

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

Loading the data and viewing a few rows

```
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 ...

head(ToothGrowth)

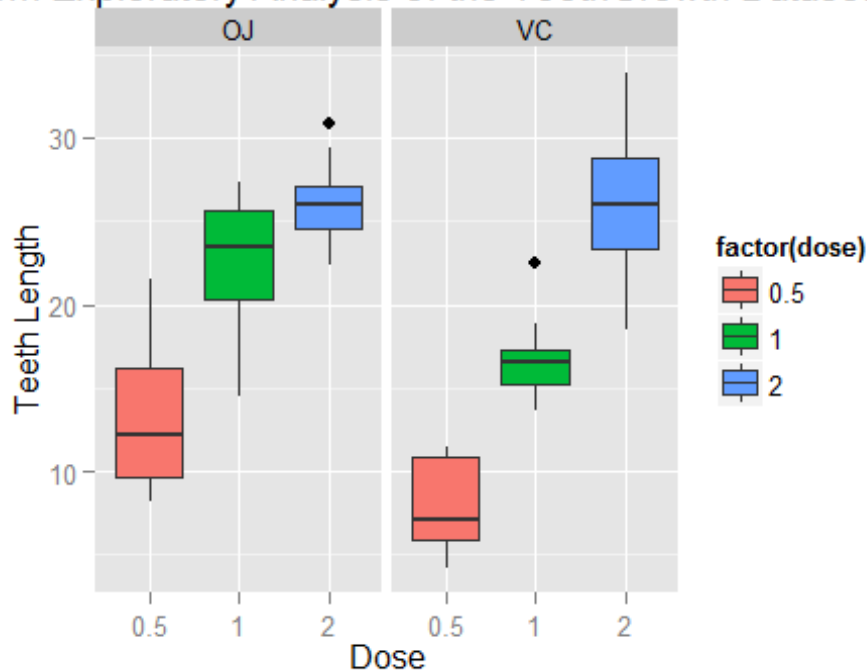
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Perform some basic exploratory data analyses

```
library(ggplot2)

ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
  ggtitle("Perform Exploratory Analysis of the ToothGrowth Dataset") +
  geom_boxplot(notch=F) +
  facet_grid(.~supp, margins = FALSE, scales = "fixed", space =
"fixed") +
  scale_x_discrete("Dose") +
  scale_y_continuous("Teeth Length")
```

orm Exploratory Analysis of the ToothGrowth Dataset



Question 2: Provide a basic summary of the data.

```
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

table(ToothGrowth$dose, ToothGrowth$supp)

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

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

```
hypo_test_1 <- t.test(len~supp, paired=FALSE, var.equal=TRUE,
data=ToothGrowth)
hypo_test_2 <- t.test(len~supp, paired=FALSE, var.equal=FALSE,
data=ToothGrowth)
```

```

hypo_test_result <- data.frame("p-value"=c(hypo_test_1$p.value,
hypo_test_2$p.value),

"Lower_Conf_Limit"=c(hypo_test_1$conf[1],hypo_test_2$conf[1]),

"Higher_Conf_Limit"=c(hypo_test_1$conf[2],hypo_test_2$conf[2]),
                        row.names=c("Equal Variance", "Unequal Variance"))

hypo_test_result

##                p.value Lower_Conf_Limit Higher_Conf_Limit
## Equal Variance  0.06039337      -0.1670064         7.567006
## Unequal Variance 0.06063451      -0.1710156         7.571016

```

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

1. Orange juice and Vitamin C have different effects on tooth growth.
2. Dosage is a key factor in tooth growth, regardless of the supplement methods.
3. Larger dosages have greater impact on tooth growth (2 mg dosage has the highest impact on tooth growth, then 1 mg. 0.5mg dosage has the lowest impact on tooth growth.
4. At a higher dosage level (2 mg), the tooth growth rate is not statistically significant for different supplement methods.