Statistical Inference Course Project 2

Xiaoyi Leo Liu

Overview

In this project, we are going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data in the R datasets package and perform some basic exploratory data analyses.
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose (only use the techniques from class, even if there's other approaches worth considering).
4. State our conclusions and the assumptions needed for our conclusions.

Load Data and Basic Summary

```
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
```

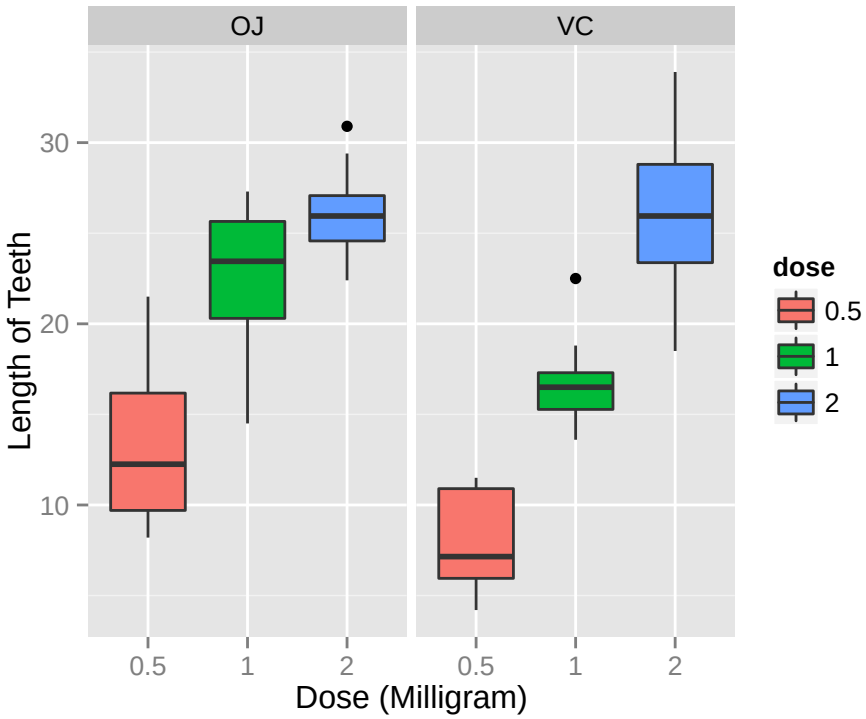
We can see that the 'dose' and 'supp' in the dataset are discrete factors.

```
ToothGrowth$dose = as.factor(ToothGrowth$dose)
ToothGrowth$supp = as.factor(ToothGrowth$supp)
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 also plot the length of teeth versus dose with different supplements.

```
library(ggplot2)
plt = ggplot(ToothGrowth, aes(x = dose, y = len, fill = dose)) + geom_boxplot(notch = FALSE) + facet_gri
plt
```



Comparing Tooth Growth

In this section, we use confidence intervals and hypothesis tests to compare tooth growth by **supp** and **dose**. The null hypothesis we assume is that the means of the two tested groups are equal, while the alternative one is the observed difference in tooth lengths is significant. We will reject the null hypothesis if the corresponding p-value is less than 0.05.

First, we will investigate the effect of supplement type on the growth of tooth for each type of dose.

```
supp_dose05 = t.test(len ~ supp, data = subset(ToothGrowth, dose = 0.5), var.equal = FALSE)
supp_dose1 = t.test(len ~ supp, data = subset(ToothGrowth, dose = 1), var.equal = FALSE)
supp_dose2 = t.test(len ~ supp, data = subset(ToothGrowth, dose = 2), var.equal = FALSE)

supp_dose = data.frame(rownames = c("dose = 0.5", "dose = 1.0", "dose = 2.0"), "p-value" = c(supp_dose05

supp_dose
```

```
##     rownames    p.value   Conf_Low Conf_High
## 1 dose = 0.5 0.06063451 -0.1710156  7.571016
## 2 dose = 1.0 0.06063451 -0.1710156  7.571016
## 3 dose = 2.0 0.06063451 -0.1710156  7.571016
```

Second, we will test the effect of dosages on tooth length.

```r
suppVC_dose05_1 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'VC' & (dose == 0.5 |dose == 1.0
suppOJ_dose05_1 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'OJ' & (dose == 0.5 |dose == 1.0

suppVC_dose05_2 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'VC' & (dose == 0.5 |dose == 2.0
suppOJ_dose05_2 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'OJ' & (dose == 0.5 |dose == 2.0

suppVC_dose1_2 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'VC' & (dose == 1.0 |dose == 2.0))
suppOJ_dose1_2 = t.test(len ~ dose, data = subset(ToothGrowth, supp == 'OJ' & (dose == 1.0 |dose == 2.0))


dose_supp = data.frame(rownames = c("supp = 'VC', dose = 0.5 or 1", "supp = 'OJ', dose = 0.5 or 1", "sup
dose_supp
```

```
##                       rownames      p.value   Conf_Low   Conf_High
## 1 supp = 'VC', dose = 0.5 or 1 6.811018e-07 -11.265712  -6.3142880
## 2 supp = 'OJ', dose = 0.5 or 1 8.784919e-05 -13.415634  -5.5243656
## 3 supp = 'VC', dose = 0.5 or 2 4.681577e-08 -21.901512 -14.4184880
## 4 supp = 'OJ', dose = 0.5 or 2 1.323784e-06 -16.335241  -9.3247594
## 5   supp = 'VC', dose = 1 or 2 9.155603e-05 -13.054267  -5.6857333
## 6   supp = 'OJ', dose = 1 or 2 3.919514e-02  -6.531443  -0.1885575
```

Conclusions

Comparing the p-value with the threshold 0.05, we can see that the null hypothesis is favored when comparing supp = "VC" and "OJ".

When we test the effect of dosages on tooth length, we find all the p-values are significantly smaller than the threshold, which means the null hypothesis should be rejected and the increase in tooth growth when supplement dose is increased is significant.