Statistical Inference Course Project: Part 2

Author: Amin Mousavi

Overview

The goal of this part of project is to analyze the ToothGrowth data set.

Load and Basic Summary of the Data

The following code is used to load the data, and summarize the basic properties of the dataset:

```
> 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 ...
```

The data has three variables len, sup and dose, and there are 60 different rows as different observations. This is the statistical summary of the variables:

```
> 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
```

Exploring the Data

To explore the data, the bar plot of tooth length with respect to dose and supp is created.

First the relation between tooth length and dose is explored by:

```
> ggplot(data=ToothGrowth, aes(x=dose, y=len, fill=supp)) +
+      geom_bar(stat="identity",) +
+      facet_grid(. ~ supp) +
+      xlab("Dose") +
+      ylab("Length")
```
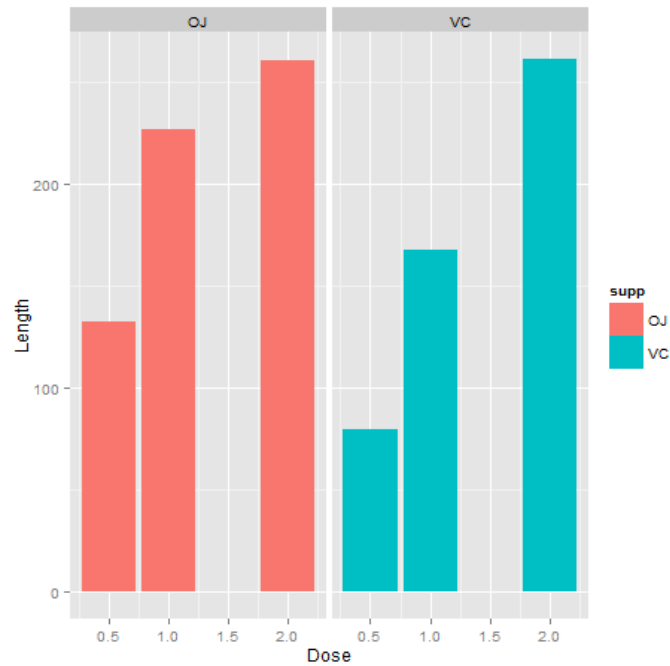
Then the relation of tooth length and supplement is investigated by this plot:

```
> ggplot(data=ToothGrowth, aes(x=supp, y=len, fill=dose)) +
+      geom_bar(stat="identity",) +
+      facet_grid(. ~ dose) +
+      xlab("Supp") +
+      ylab("Length")
```
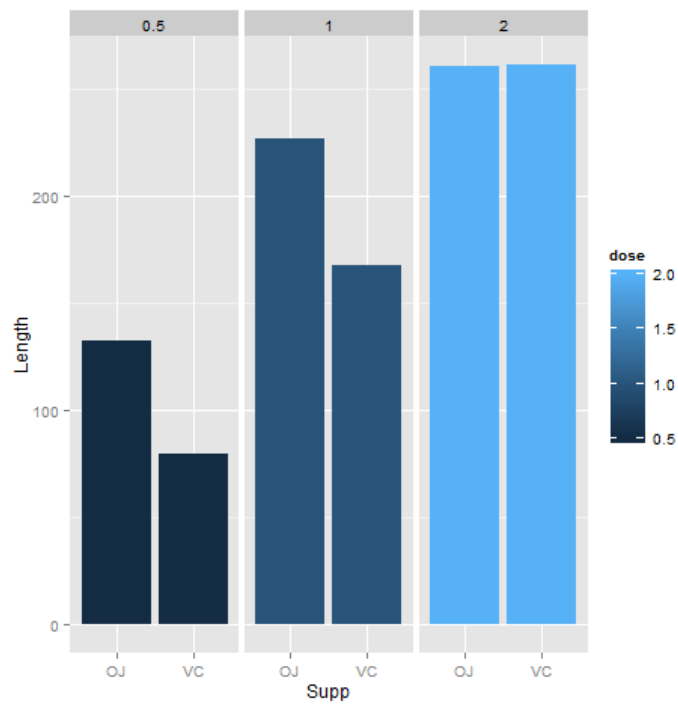
It seems that there is a negative correlation between tooth length and supplement.

Compare Tooth Growth by Dose and Supplement

To study the relation between tooth length and dose and sup, we can use t hypothesis test:

```
> tests = list()
> dose = c(0.5,1,2)
> for (d in dose) {
+     ojd = ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "OJ"]
+     vcd = ToothGrowth$len[ToothGrowth$dose == d & ToothGrowth$supp == "VC"]
+     t <- t.test(ojd, vcd)
+     id <- paste0("OJ","-", "VC",",",d)
+     tests <- rbind(tests, list(id=id, p.value=t$p.value, conf.lo=t$conf.int
[1], conf.hi=t$conf.int[2]))
+ }
> tests
     id          p.value      conf.lo   conf.hi
[1,] "OJ-VC,0.5" 0.006358607 1.719057 8.780943
[2,] "OJ-VC,1"   0.001038376 2.802148 9.057852
[3,] "OJ-VC,2"   0.9638516    -3.79807 3.63807
```

Conclusion

From the t test result, we can conclude that:

We cannot claim any assumption about the effect of OJ and VC in dose 2. Therefore, we cannot conclude that in dose 0.5 and dose 1, the length with OJ is longer than VC.