Statistical Inference Course Project Part 2

Aayush Shah

Overview

In this project, a dataset called *ToothGrowth* is being investigated with exploratory data analysis. The *ToolthGrowth* dataset contains observation for the effect of vitamin C on tooth growth in Guinea Pigs, which corresponds to the response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Exploring the *ToothGrowth* 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 ...
```

```
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
```

```
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
```

The Summary tells us that 30 observations are related to orange juice (supp = "OJ"), and other 30 are with ascorbic acid (supp = "VC"). The *len* (length) observation has its range as (4.20,33.90) with mean = 18.81 and median = 19.25.

Boxplot

Lets use a simple boxplot to visualize what we see in the summary.

```
library(ggplot2)
ggplot(ToothGrowth, aes(x=dose, y=len)) + geom_boxplot(aes(fill=factor(dose))) + facet_grid(.~supp) + g
```
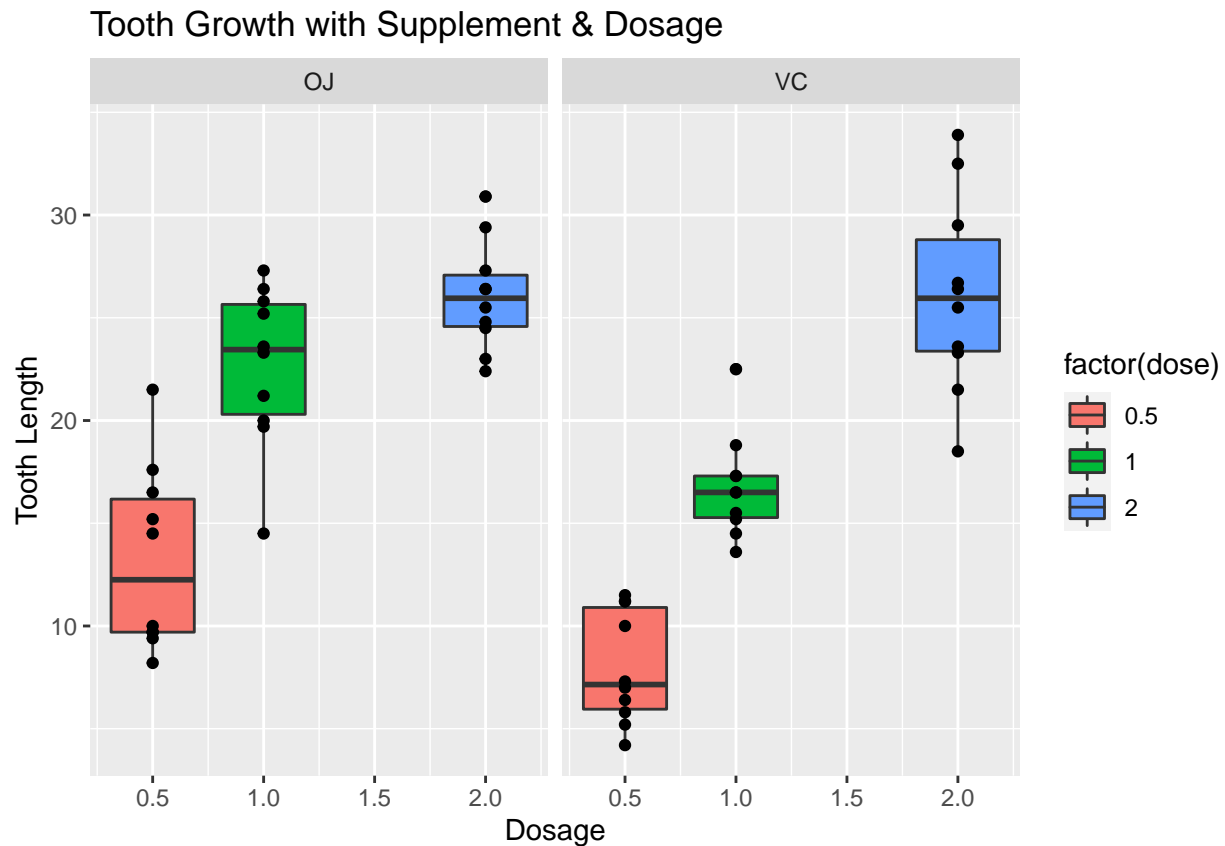

So from the graph, it is clear that when the dosage is at 2 mg, the mean value of tooth growth is similar in *OJ(Orange juice)* and *VC(Vitamin C)* supplements. But, when the dosage is 0.5 mg or 1 mg, the boxplot shows that OJ has a more positive impact on tooth growth compared to VC.

Hypothesis Testing

Say,

H0: *Orange juice(OJ)* has a **more positive** impact on tooth growth than *Ascorbic Acid(VC)*.

H1: *Orange juice(OJ)* does **not** have a **more positve** impact on tooth growth than *Ascorbic Acid(VC)*.

```
t.test(len ~ supp, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
```

```
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##          20.66333         16.96333
```

```
t.test(len ~ supp, data=ToothGrowth[ToothGrowth$dose < 2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##            17.965           12.375
```

```
t.test(len ~ supp, data=ToothGrowth[ToothGrowth$dose == 2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##             26.06            26.14
```

At 95% confidence interval (CI), the first two tests do not include 0. Hence there is a significant difference between the supplementing with OJ and VC. However, the difference diminishes when the dosage is above 2 mg.

Conclusion

Assuming the supplements, orange juice and ascorbic acid were iid among the subjects, it can be concluded that when OJ is supplied with dosage *less than 2 mg* it has a significant positive impact on tooth growth.