Statistical Inference Course Project Part 2

Ankitha Giridhar

18/10/2021

Overview

The second part of the Statistical Inference course project is explored here. A basic exploratory data analysis is done on the ToothGrowth data, followed by hypothesis testing. The conclusions from the latter are presented at the end.

EDA

The data is examined using the summary() and head() functions.

```
data("ToothGrowth")
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
```

There are only three columns in the data - the length, the supplement and the dose. The dose and supplement columns are further looked into for unique values.

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
unique(ToothGrowth$supp)
```

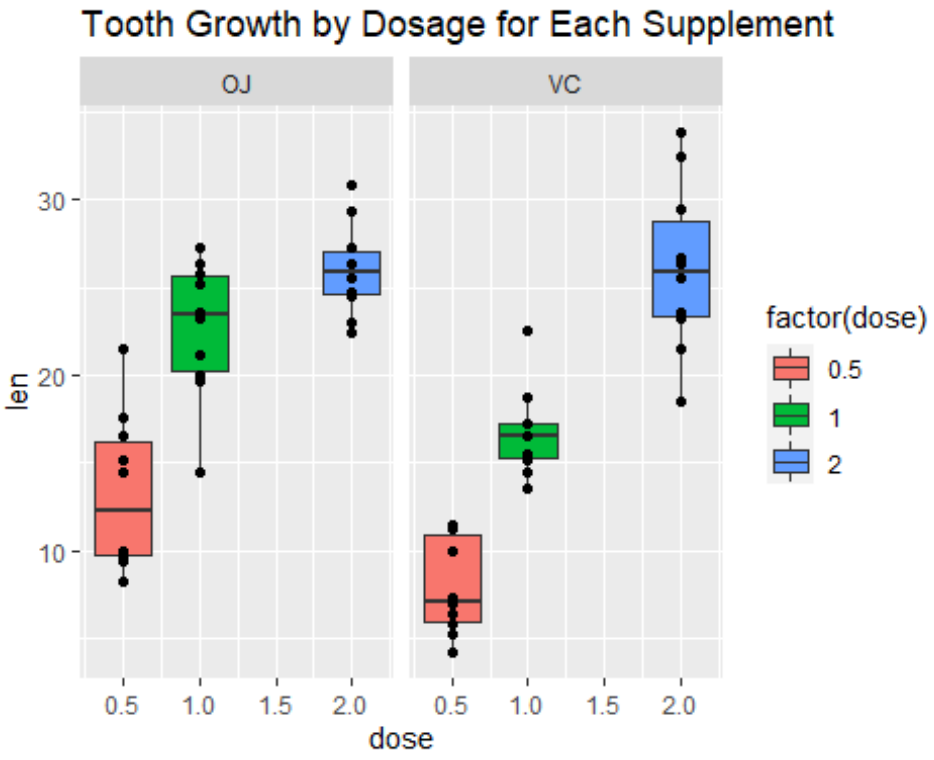
```
## [1] VC OJ
## Levels: OJ VC
```

There are three unique dose values, while the supplement column is binary. The data is visualised on these categories using the ggplot library.

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

p <- ggplot(ToothGrowth, aes(x=dose, y=len)) +
geom_boxplot(aes(fill=factor(dose))) + geom_point() + facet_grid(.~supp) +
ggtitle("Tooth Growth by Dosage for Each Supplement")
p
```



In both cases, higher doses correlate with greater lenggth. For each dosage, OJ seems to have higher lengths than VC.

Hypothesis Testing

The hypotheses are that orange juice (OJ) has a higher impact than ascorbic acid (VC) and vice versa. These are tested with the t test.

```
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
```

The p value was greater than 0.5, and this test is inconclusive, as the null hypothesis cannot be rejected. Moreover, the confidence interval includes 0. The test is tried again with doses less than 2, the highest dose in the data.

```
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
```

Here, the the p value is ~0.004, i.e. close to 0, and the confidence interval does not include 0. ### Conclusion Therefore, it can be concluded that for dosages less than, NOT equal to 2, the supplement has an effect on the tooth growth of guinea pigs, with orange juice resulting in longer teeth as per the means.