Statistical Inference Part - 2

Basic inferential data analysis

Niraj Nair

6 June, 2020

Assignment

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data 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 your conclusions and the assumptions needed for your conclusions.

Required Packages

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

1. Loading the ToothGrowth data and performing some basic exploratory data analyses.

```r
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 ...
```

```r
head(ToothGrowth)
```

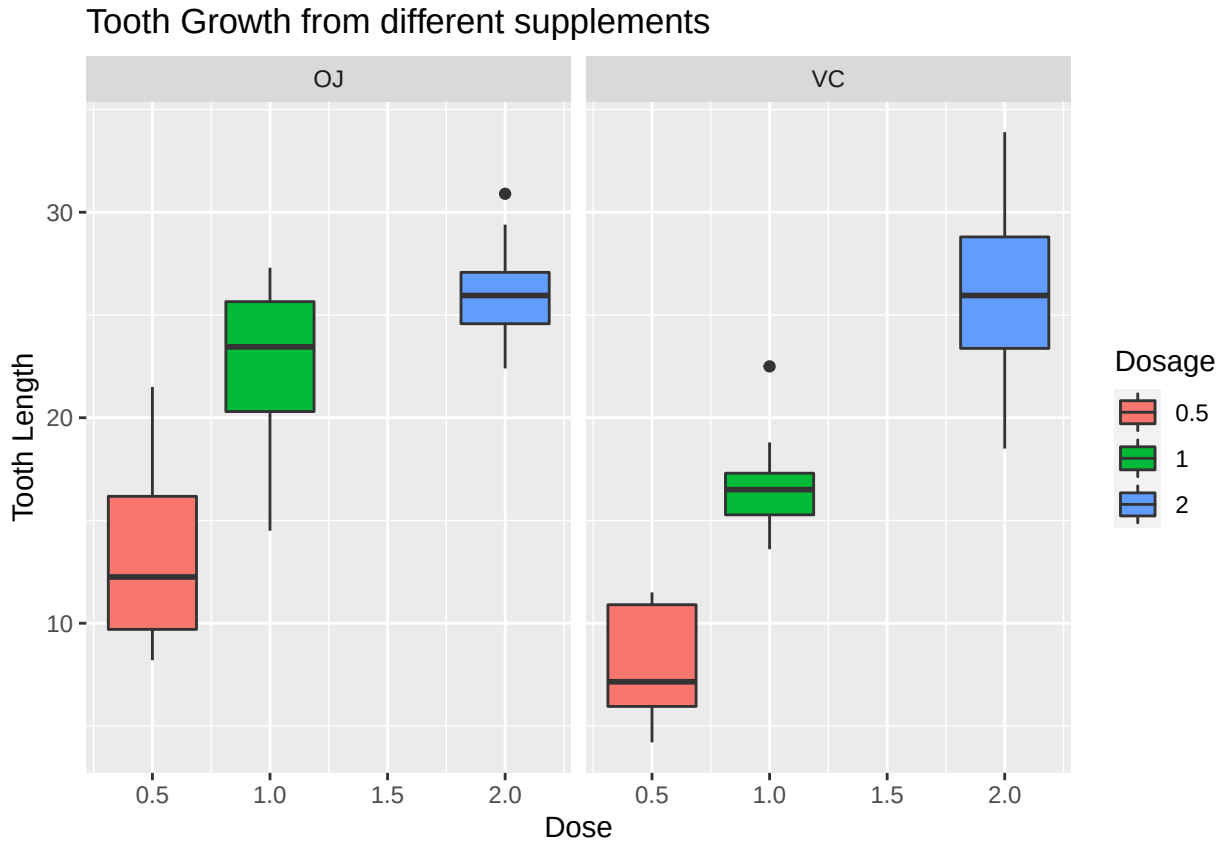
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

2. Obtaining a summary of the dataset.

```r
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
```

```r
g <- ggplot(data = ToothGrowth, aes(x = dose, y = len, fill = factor(dose)))
g + geom_boxplot()+
    facet_grid(.~supp)+
    labs(title = "Tooth Growth from different supplements", x = "Dose", y = "Tooth Length") +
    scale_fill_discrete(name = "Dosage")
```




```
growth <- ToothGrowth %>% group_by(supp, dose) %>% summarise(len = mean(len))
growth
```

```
## # A tibble: 6 x 3
## # Groups:   supp [2]
##   supp   dose   len
##   <fct> <dbl> <dbl>
## 1 OJ      0.5 13.2
## 2 OJ      1   22.7
## 3 OJ      2   26.1
## 4 VC      0.5  7.98
## 5 VC      1   16.8
## 6 VC      2   26.1
```

3. Using confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

```
OJ <- ToothGrowth$len[ToothGrowth$supp == "OJ"]
VC <- ToothGrowth$len[ToothGrowth$supp == "VC"]
t.test(OJ, VC, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = .95)
```

(a) We assume the null hypothesis to be that there is no relation betweem tooth growth and supp.

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Since the P-value is 3% ie. $< 5\%$, we can reject the null hypothesis. This conveys that the alternative hypohtesis is accepted and therefore there exists a relation between tooth lenght and supp.

```
d05 <- ToothGrowth$len[ToothGrowth$dose == .5]
d1 <- ToothGrowth$len[ToothGrowth$dose == 1]
d2 <- ToothGrowth$len[ToothGrowth$dose == 2]
```

(b) We assume the null hypothesis to be that there is no relation betweem tooth growth and dose.

- First we will check that if the alternative hypothesis is right for the smaller doses

```
t.test(d05, d1, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = .95)
```

```
##
##  Welch Two Sample t-test
##
## data:  d05 and d1
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

- Second we will check that if the alternative hypothesis is right for the bigger doses

```
t.test(d1, d2, alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = .95)
```

```
##
##  Welch Two Sample t-test
##
## data:  d1 and d2
## t = -4.9005, df = 37.101, p-value = 9.532e-06
```

```
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

From both the tests we can clearly see that the P-values are very low. Therefore the null hypothesis is rejected and we can say that the tooth growth depends on the doasage of supp.

Conclusion

- There doesn't seem to be a statistically significant difference between delivery methods, with OJ apparently more effective at dose levels 0.5 and 1, and VC slightly more effective at dose level 2.
- It appears that there is a statistically significant difference between tooth length and dose levels across both delivery methods, in other words, as the dose increases so does tooth length.

‘‘‘