Statistical Inference Project 2

by,

Ravi Teja L

22-06-2020

Part-2

Examining ToothGrowth In R

Instructions

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.

Loading Libraries

```
library("data.table")
library("ggplot2")
library("dplyr")
```

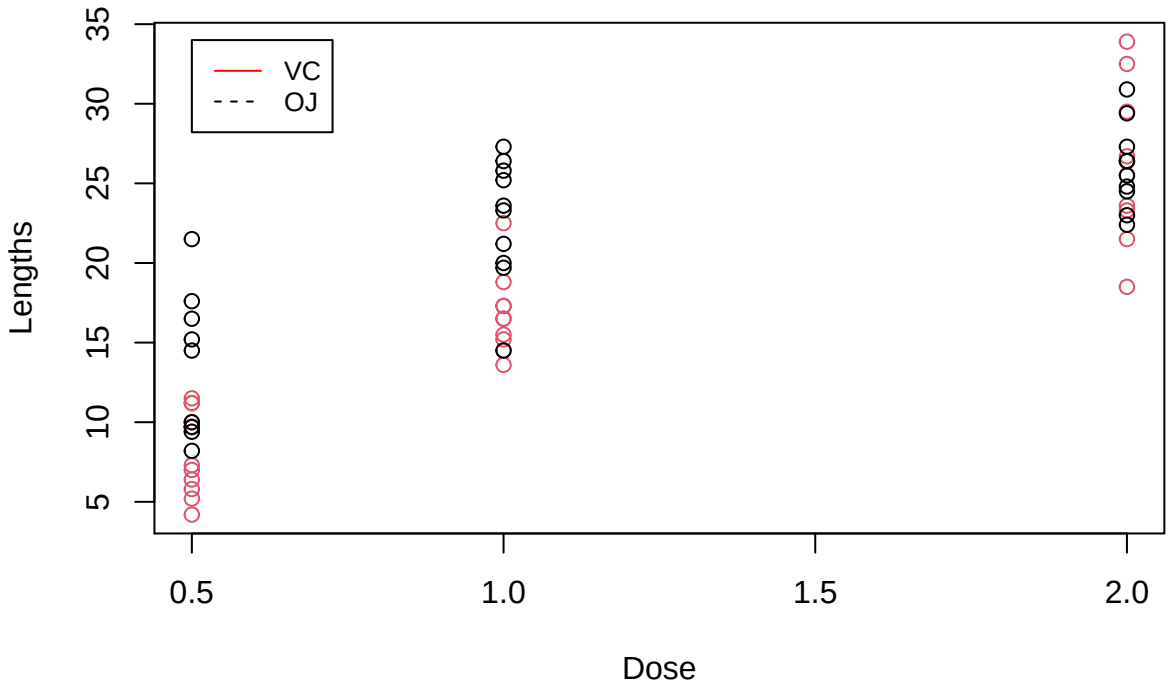
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

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

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

```
library(tidyr)
library(ggpubr)
```

Load the ToothGrowth data and perform some basic exploratory data analyses

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

```
ToothGrowth <- data.table(ToothGrowth)
plot(x = ToothGrowth$dose, y = ToothGrowth$len, xlab = "Dose", ylab = "Lengths", col = ToothGrowth$supp
legend(0.5, 34, legend=unique(ToothGrowth$supp), col = c("red", "black"), lty=1:2, cex=0.8)
```



From the graph, it can be observed that as the dosage increases tooth length increases.But it also appears that OJ tooth delivers more tooth growth even at lower dosages

Basic summary of Data

```
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
```

Examining Dose and Supplement

```
ToothGrowth[, .(meanlen = mean(len), Count = .N), by = .(dose,supp)]
```

```
##    dose supp meanlen Count
## 1:  0.5   VC    7.98    10
## 2:  1.0   VC   16.77    10
## 3:  2.0   VC   26.14    10
## 4:  0.5   OJ   13.23    10
## 5:  1.0   OJ   22.70    10
## 6:  2.0   OJ   26.06    10
```

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

By the summary the main relation camparision is between the usages of supplement to tooth lengths. So the tests are to be made on supplements and tooth length for different type of dosages

```
lowdose <- ToothGrowth[ToothGrowth[, dose == 0.5]]
middose <- ToothGrowth[ToothGrowth[, dose == 1.0]]
highdose <- ToothGrowth[ToothGrowth[, dose == 2.0]]
```

```
t.test(ToothGrowth$len ~ ToothGrowth$supp)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len by ToothGrowth$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 above test does show a wide confidence interval but the significance of p-value shows higher value so we can conclude and donot reject Null Hypothesis.

```
t.test(lowdose$len ~ lowdose$supp, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  lowdose$len by lowdose$supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

From the above test we can find a lot of variation between means of two supplements and low p-value. That definitely concludes the fact that OJ shows better results that VC

```
t.test(middose$len ~ middose$supp, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  middose$len by middose$supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

From the above test we can find a lot of variation between means of two supplements and low p-value. That definitely concludes the fact that OJ shows better results that VC

```
t.test(highdose$len ~ highdose$supp, paired = FALSE)
```

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

From the above test we can find very low variation between two supplements. Here we can conclude that OJ and VC show similar results in higher dosages

Conclusions and Assumptions

Based off this data we can conclude the following:

As the dosage increases, tooth length increases regardless of what kind of supplement we use. At the 0.5 mg and 1.0 mg dosage the OJ supplement method provides more tooth growth than that of VC Supplement At the 2.0 mg dosage, there is no significant difference between the OJ and VC supplement methods.

Assumptions

We assume that the measurements are not paired. We assume the populations are independent, that there was no crossover between the subjects and dosage.

On an overall scenario we can consider that OJ performs better than VC in most of the cases