Analysis of ToothGrowth Data from datasets package in R

Vedant Mane

June 16, 2020

In this project we will perform analysis on the ToothGrowth data in the R datasets package and perform some hypothesis tests on the data in the context of finding the relations within the several attributes of the dataset.

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- 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)
- State your conclusions and the assumptions needed for your conclusions.

Loading Libraries & Data

```
#Loading essential Libraries
library(datasets)
library(ggplot2)
#Loading ToothGrowth data and looking at its contents
data(ToothGrowth)
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
```

```
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 ...
```

Performing Exploratory Data Analysis

Performing some Exploratory Data Analysis in order to get an idea about the contents of the Dataset like for example how data is distributed based on different supplements & amount of doses.

```r
#Plotting a graph using ggplot function to visualize the relations
#between various attributes in the Dataset
ggplot(data = ToothGrowth, aes(dose, len, color = supp)) +
     geom_point() +
     geom_smooth(method = "lm")
```

`geom_smooth()` using formula 'y ~ x'
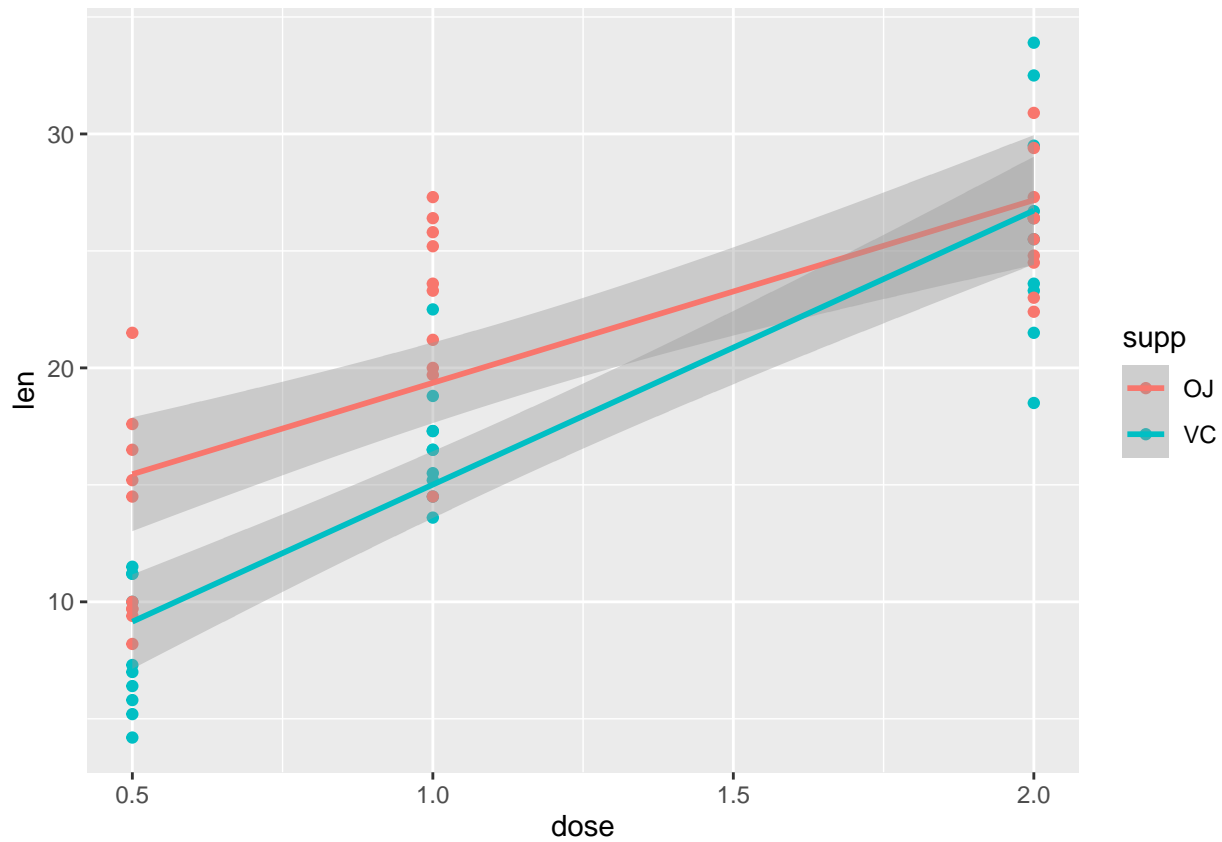
Summarizing the Dataset

Summarinzing the Dataset to find out additional information about the extent of various attributes of the data in the ToothGrouwth Dataset.

Categorizing based on levels of the Supplements and Amount of Doses attributes to analyze their results on the change in the Tooth Length based on the medication for several patients.
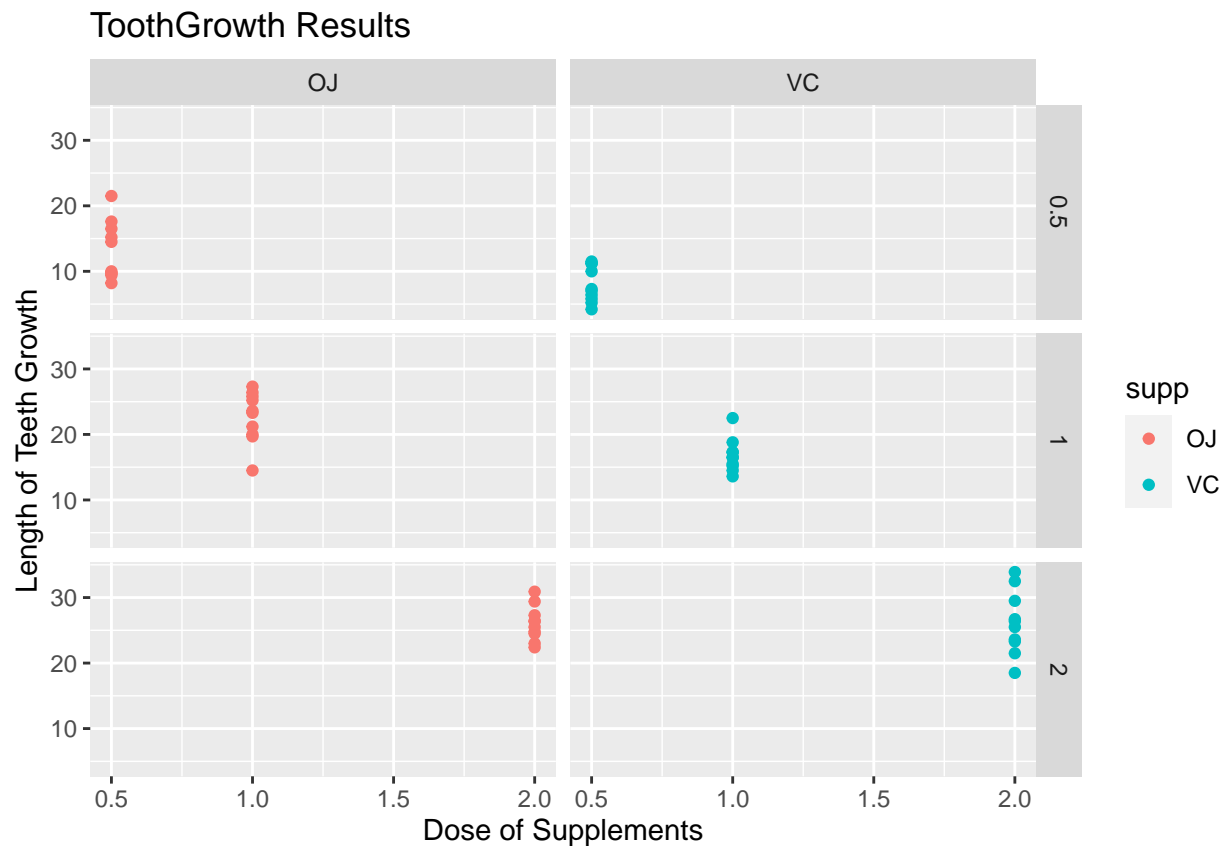
```r
summary(ToothGrowth)
```

```
##       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
```

```
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

```
ggplot(ToothGrowth, aes(dose, len, color = supp)) +
    geom_point() + facet_grid(dose ~ supp) +
    ggtitle("ToothGrowth Results") +
    xlab("Dose of Supplements") + ylab("Length of Teeth Growth")
```


Hypothesis Test - Based on Type of Supplement

Now, let us test our data to see the relation based on the type of supplements given to the patients.

```
suppTest <-t.test(len ~ supp, data = ToothGrowth, paired = FALSE)
```

We perform the Welsh Two Sided Test for a 95% confidence interval and it can be seen that it has the range of **-0.1710156, 7.5710156**.

We see that our range consists of Zero and hence we cannot conclude that one supplement is better than the other from our data because this may be the cause of the amount of doses given to the patients.

Hypothesis Test - Based on Amount of Doses

So, lets test our data for the amount of doses given to the patients and observe the results.

```
subData1 <- ToothGrowth[ToothGrowth$dose %in% c(0.5,1),]
doseTest1 <- t.test(len ~ dose, data = subData1, paired = FALSE)
subData2 <- ToothGrowth[ToothGrowth$dose %in% c(0.5,2),]
doseTest2 <- t.test(len ~ dose, data = subData2, paired = FALSE)
subData3 <- ToothGrowth[ToothGrowth$dose %in% c(1,2),]
doseTest3 <- t.test(len ~ dose, data = subData3, paired = FALSE)
```

We perform the Welsh Two Sided Test for a 95% confidence interval for the doses 0.5 & 1 and it can be seen that it has the range of **-11.9837813, -6.2762187**

We perform the Welsh Two Sided Test for a 95% confidence interval for the doses 0.5 & 2 and it can be seen that it has the range of **-18.1561665, -12.8338335**

We perform the Welsh Two Sided Test for a 95% confidence interval for the doses 1 & 2 and it can be seen that it has the range of **-8.9964805, -3.7335195**

From the above results, this is the case where the amount of doses given to the patients is directly proportional to the growth of Teeth over a specific period of time.

Conclusion

Taking into account the results of the summary, analysis of the data and the various hypothesis tests performed we are able to conclude that the type of supplement does not play a major role in getting the results which in our case is increasing the length of teeth but it can be seen that the amount of doses have a crucial role in getting the positive results. This concluion supports our initial exploratory data analysis which is visualized in the graph.