Exploratory Analysis of ToothGrowth dataset in R

Yanal Kashou

Contents

Overview of Assignment 1
Data Loading 1
Data Summary 1
Analysis 2
T-test Inference 2
Assumptions 3
Conclusions 3

Overview of Assignment

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.

Data Loading

```
# Load necessary libraries
library(datasets)
library(ggplot2)
library(Hmisc)
```

Data Summary

```
# Load and explore dataset
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 ...
```

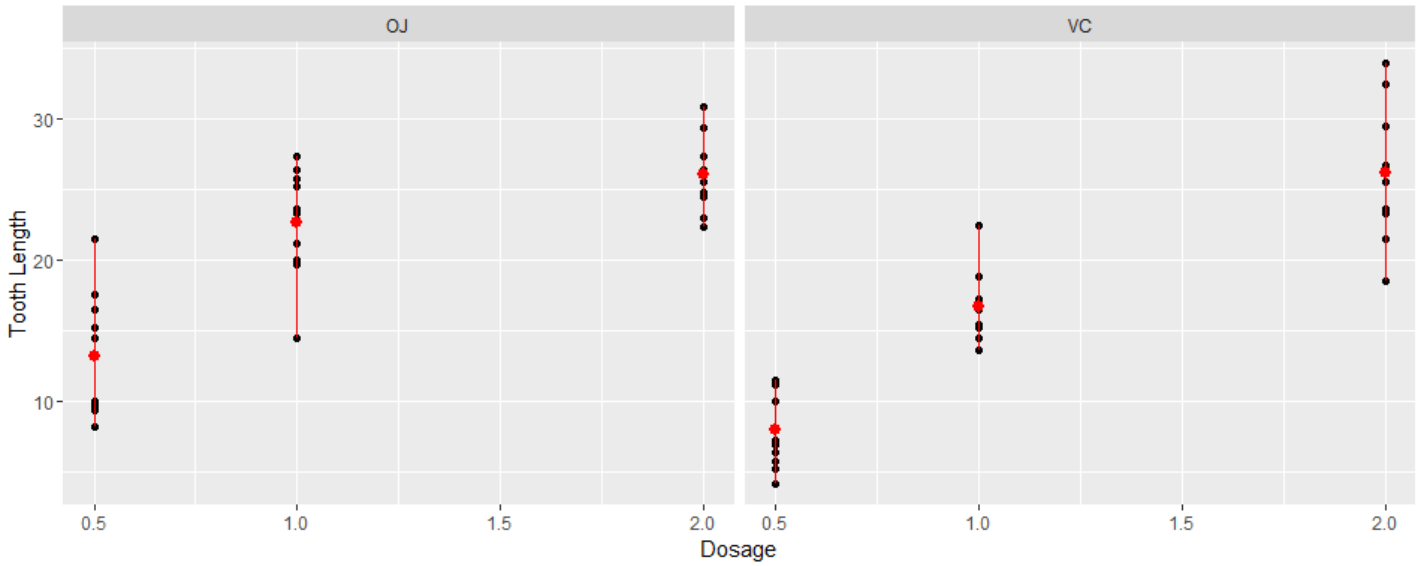
```
head(ToothGrowth, 10)
```

```
##     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
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
```

```
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

Analysis

```
# Plot Dosage Vs Tooth Length
d <- ggplot(ToothGrowth, aes(dose, len)) + geom_point() + facet_grid(. ~ supp) +
    xlab("Dosage") + ylab("Tooth Length")
d + stat_summary(fun.y = mean, fun.ymin = min, fun.ymax = max, colour = "red")
```



T-test Inference

```
# T-tests at dosage levels 0.5, 1 and 2
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 0.5, ])

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

t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])

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

```
t.test(len ~ supp, 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
```

Assumptions

The only assumption is that orange juice will yield better tooth growth than pure Vitamin C due to the presence of other nutrients, vitamins, and minerals.

However this is also a conclusion inferred from our analysis.

Conclusions

We applied T-tests at all three dosage levels and we can safely conclude:

1. Higher dosage corresponds to higher effect (within range)
2. Both Vitamin C and Orange Juice aided in tooth growth
3. Orange Juice gave better results probably due to other unidentified present nutrients.