The Influence of Vitamin C on Guinea Pigs' Tooth Growth

Valerii Podymov

October 24, 2015

Overview

The goal of this project is to analyze the ToothGrowth data in the R `datasets` package. The data is the result of measuring the effect of different dosage of Vitamin C on the length of teeth of 60 guinea pigs. Vitamin C was supplemented with Orange Juice and Ascorbic Acid in three different dosage, 0.5, 1.0, and 2.0 milligrams/day. The focus of the report is statistical inference around different categorizations on the length of teeth.

Data Loading, Cleaning and Exploratory Analysis

We start with loading data from the source and performing some basic exploratory data analyses.

```
# Load necessary libraries
library(plyr)
library(lattice)
library(datasets)
```

```
# Load ToothGrowth data
dt <- ToothGrowth
```

To understand the content and the structure of the data we use the following two methods.

```
# Provide a basic summary of the data
summary(dt)
```

```
##       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
```

```
# Display the internal structure of the data.frame object
str(dt)
```

```
## '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 ...
```

Now we check if the data contain missing values.

```r
sum(is.na(dt))
```

```
## [1] 0
```

Since there are no missing values, we can continue the analysis. Next step is to change the names of variables to something more meaningful.

```r
colnames(dt) <- c("Length", "Supplement", "Dosage")
```

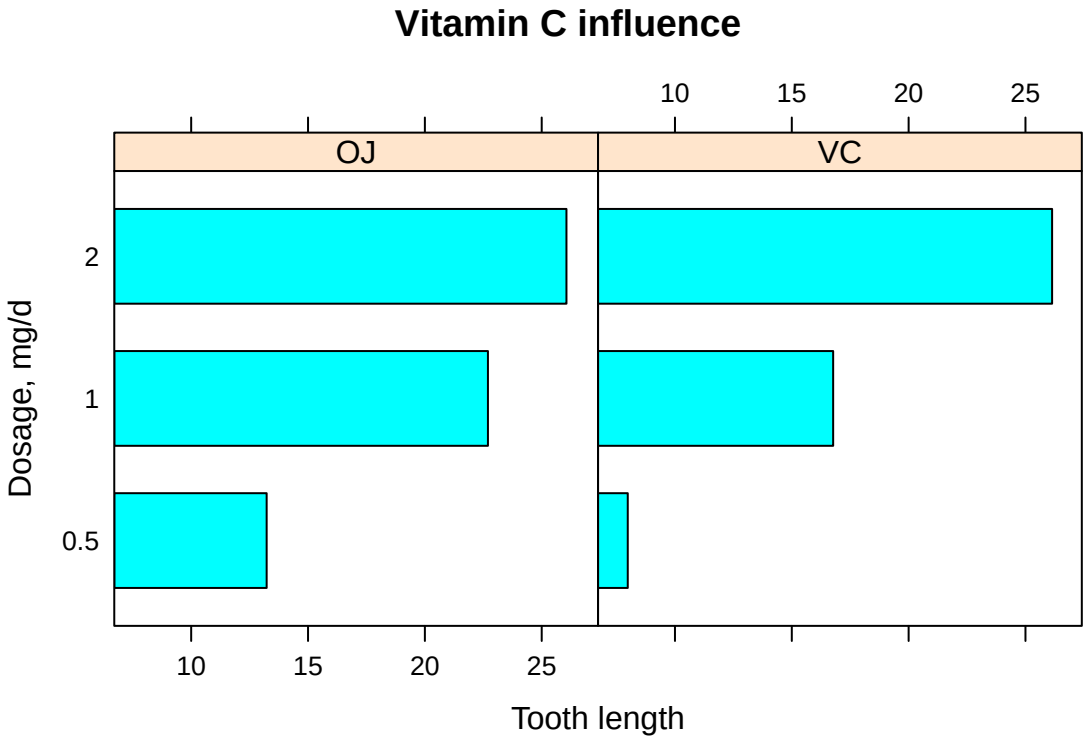
For further analysis we need to visualize tooth lenght against the dosage and supplement. For this purpose we first split data into subsets by

- dosages - 0.5, 1.0 and 2.0 mg/d
- supplements - orange juice (`OJ`) and ascorbic acid (`VC`)

and compute the mean average for each.

```r
# Compute summary statistics of data subsets
avg_length <- aggregate(Length ~ Dosage + Supplement, dt, mean)
```

The following chart shows that there is a strong relationship between the average teeth length and dosage.



However, it is still unclear which supplement, Orange Juice `OJ` or Ascorbic Acid `VC`, has more influence on teeth growth.

Confidence Interval Testing

In order to understand Vitamin C effect on the teeth growth, we conduct the following confidence interval testing scenarios

- Dependence of the teeth growth on the dosage
- Dependence of the teeth growth on the supplement
- Dependence of the teeth growth on different combinations of two factors

in the followng assumptions

- the data samples are independent and normally distributed
- the groups received orange juice and ascorbic acid are independent.

For each of the scenarios we subset available data appropriately and utilize the `t.test` function to determine confidence intervals.

Dosage

Let *H0* be a null hypotesis

H0: the increase of the dosage does not result in the average tooth length growth

```
# Subsetting the data set
t1 <- subset(dt, Dosage == 0.5)$Length
t2 <- subset(dt,Dosage==1.0)$Length
# Perform t-confidence interval test
t <- t.test(t1, t2, paired = FALSE, var.equal = FALSE)
t$conf.int[1:2]
```

```
## [1] -11.983781  -6.276219
```

Since the confidence interval does not contain zero we can reject the null hypothesis. In the other words, there is a positive correlation between the increase of the dosage from 0.5 to 1.0 mg/d and the average teeth length growth.

Next, if we increase the Vitamin C dosage from 1.0 to 2.0 milligrams per day, the confidence interval also does not contain zero.

```
t1 <- subset(dt, Dosage == 1.0)$Length
t2 <- subset(dt, Dosage == 2.0)$Length
t <- t.test(t1, t2, paired = FALSE, var.equal = FALSE)
t$conf.int[1:2]
```

```
## [1] -8.996481 -3.733519
```

Consequently, we should reject the null hypothesis again.

Type of Supplement

Let *H0* be a null hypotesis

H0: the type of the supplement does not have an influence on the average tooth length growth

```r
t1 <- subset(dt, Supplement == 'VC')$Length
t2 <- subset(dt, Supplement == 'OJ')$Length
t <- t.test(t1, t2, paired = FALSE, var.equal = FALSE)
t$p.value
```

```
## [1] 0.06063451
```

```r
t$conf.int[1:2]
```

```
## [1] -7.5710156  0.1710156
```

In this single test the p-value is 0.0606345 and the confidence interval contains zero, so we do not reject the null hypothesis and conclude that the type of Vitamin C supplement alone does not affect the average teeth length.

Supplement and Dosage in Combinations

Now we compare the influence of each of the dosage of orange juice with equivalent dosage of ascorbic acid.

Let null hypotesis be

H0: the type of the supplement with the dosage of 0.5 mg/d does not have an influence on the average tooth length growth

```r
t1 <- subset(dt, Supplement == 'VC' & Dosage == 0.5)$Length
t2 <- subset(dt, Supplement == 'OJ' & Dosage == 0.5)$Length
t <- t.test(t1, t2, paired = FALSE, var.equal = FALSE)
t$conf.int[1:2]
```

```
## [1] -8.780943 -1.719057
```

The confidence interval does not contain zero, so we should reject the null hypothesis.

Then we check the next null hypotesis

H0: the type of the supplement with the dosage of 1.0 mg/d does not have an influence on the average tooth length growth

```r
t1 <- subset(dt, Supplement == 'VC' & Dosage == 1.0)$Length
t2 <- subset(dt, Supplement == 'OJ' & Dosage == 1.0)$Length
t <- t.test(t1, t2, paired = FALSE, var.equal = FALSE)
t$conf.int[1:2]
```

```
## [1] -9.057852 -2.802148
```

The confidence interval does not contain zero, so we should reject the null hypothesis.

And finally we check the null hypotesis

H0: the type of the supplement with the dosage of 2.0 mg/d does not have an influence on the average tooth length growth

```r
t1<-subset(dt,Supplement=='VC' & Dosage==2.0)$Length
t2<-subset(dt,Supplement=='OJ' & Dosage==2.0)$Length
t<-t.test(t1,t2,paired=FALSE,var.equal=FALSE)
t$p.value
```

```
## [1] 0.9638516
```

```r
t$conf.int[1:2]
```

```
## [1] -3.63807  3.79807
```

The confidence interval contains zero and the p-value is almost 1. Thus, we do not reject the null hypothesis and it means that with the dosage of 2.0 mg/d we cannot conclude which supplement type has a greater effect on teeth growth.

Conclusions

- As Vitamin C dosage alone increases, the average teeth length increases as well
- Irrespective of the dosage, the supplement type alone does not affect the teeth growth
- Orange Juice affects the average teeth length greater then Ascorbic Acid with the dosage of 0.5 and 1.0 mg/d
- When the dosage reaches the value of 2.0 mg/d, there is no significant difference between Orange Juice and Ascorbic Acid