Hypothesis Testing

Pedro A. Alonso Baigorri

11 de octubre de 2017

Synopsis

In this document I'm going to analyze the ToothGrowth data in the R datasets package with the objective to perform and hypothesis tests to compare tooth growth by supp and dose.

The ToothGrowth data, provides data samples about the Effect of Vitamin C on Tooth Growth in Guinea Pigs.

Summary of the data

First of all, I'll load the data and I will perform some basic analysis over the data.

```
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
```

```
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
```

It can be show that the dataset contains 60 observations with 3 variables:

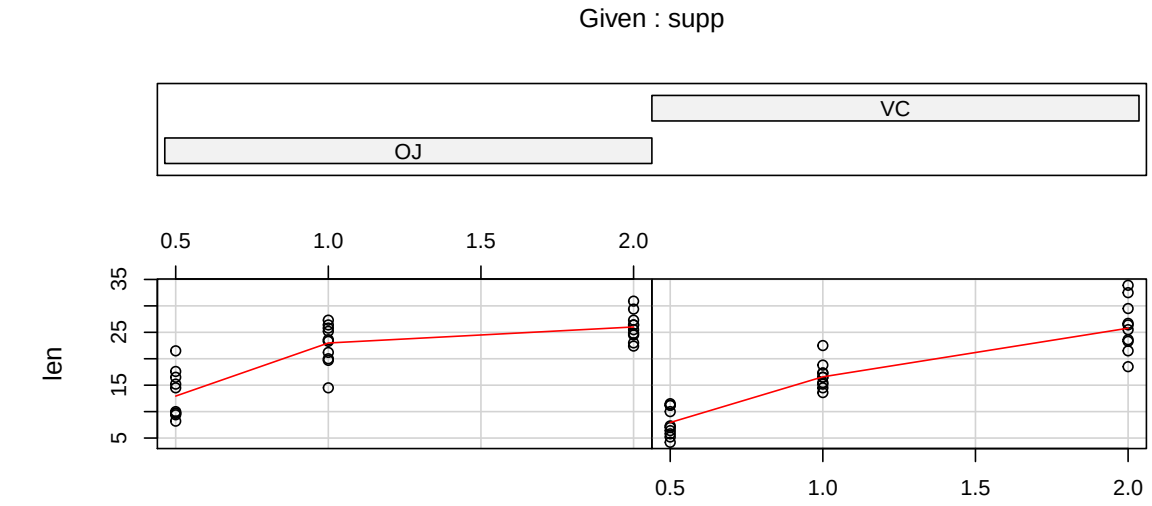
- len: tooth length
- supp: the sumplement type
- dose: the doses in mgms/day

The number of measures by the different levels of dose and supp are identical as can be shown now:

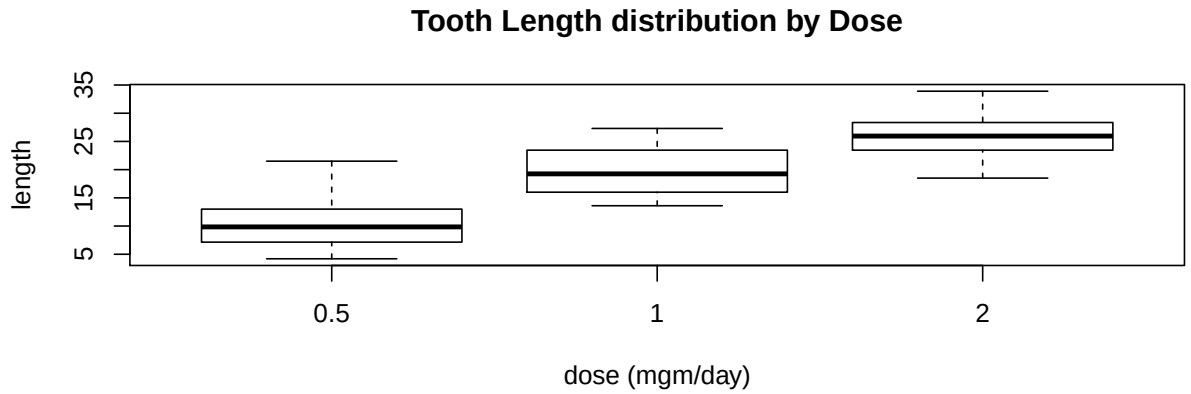
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

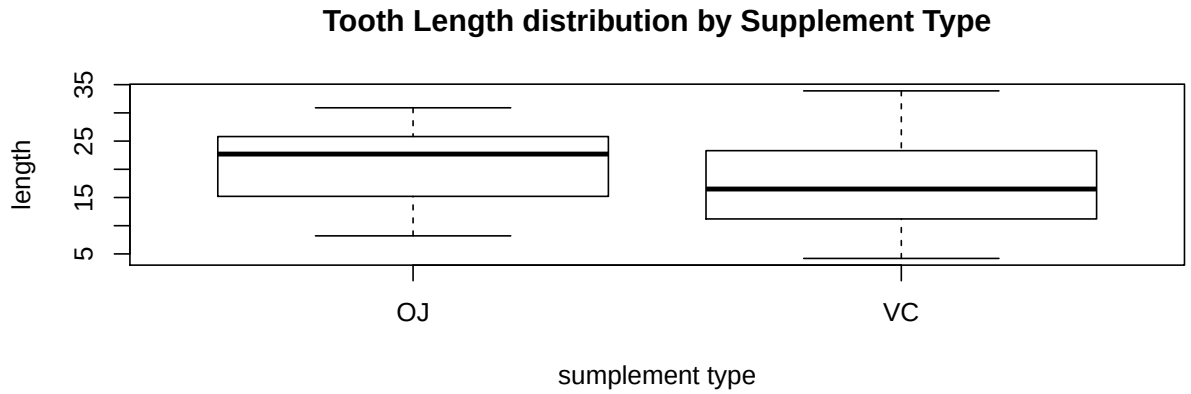
```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

The following plots gives more information about the dataset:



ToothGrowth data: length vs dose, given type of supplement





According to the plots we can see how the length is increasing with the number of dosis. And in general it seems that the length is also higher with the supplement type is OJ.

Assumptions

To perform the hypothesis tests I will do the following assumptions:

- The population for the experiment follows a normal distribution
- The variance of the population for different supplement types are similar.

Hypothesis tests

I will test the Hypothesis Ho that the mean for different type of supplement type (OJ | VC) are identical. This is that the differences of the mean = 0. So Ho (mean(x|OJ) - mean (x|VC) = 0).

Then the althernative hypothesis is Ha (mean(x|OJ) - mean (x|VC) =! 0).

In the first test I will use all the samples including all the doses quantities. For all the tests I will use alpha = 0.1.

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  0.4708204 6.9291796
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

According to this results the p-value (0.0603934) is lower then 0.1 so we reject the null hypothesis.

This can be confirmed seeing the confidence intervals for the difference of the mean:

0.4708204, 6.9291796

than are clearly above 0.

Now, I will perform a new set of hypothesis tests but now I will separate in different tests the samples for each level of doses to see the effect of the doses in the experiment.

```
## Dose =  0.5 ; p-value: 0.005303661 Confidence intervals: 2.377886 8.122114
## Dose =  1 ; p-value: 0.0007807262 Confidence intervals: 3.38014 8.47986
## Dose =  2 ; p-value: 0.9637098 Confidence intervals: -3.086866 2.926866
```

In this case we can see that for the first 2 levels of doses (0.5 and 1) we can reject the null hypothesis as the p.values are very low and confidence intervals are always above 0.

But in the case of dose = 2, we can't reject Ho, because p-value is higher than alpha. We can see also that in this case the mean difference = 0 is inside of the confidence interval.

Conclusions

According to the tests we can conclude:

- For the doses = 0.5 & 1, there are a significance difference between the samples with the different types of Supplement Types: OJ Vs VC
- However, when the dose = 2, we can't say that there is a significance difference so we can't conclude that the different Supplement Types affect to the increase of the tooth length.

Annexes

Code for Hypothesis tests

```
alpha <- 0.10

# for all doses
t <- t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = ToothGrowth,
    alternative = "two.sided",conf.level = 1-alpha, mu = 0)
t
```

```
ToothGrowth$dosef <- as.factor(ToothGrowth$dose)

# different test for each levels of doses
for (i in levels(ToothGrowth$dosef))
{
    s <- subset(ToothGrowth, dosef == i)
    test <- t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = s,
                   alternative = "two.sided",conf.level = 1-alpha, mu = 0)

    cat("Dose = ",i,"; p-value:",test$p.value, "Confidence intervals:",
            test$conf.int, "\n")
}
```