

Statistical Inference - Project Assignment 2

Philip Ohlsson

Overview

This project will analyze the ToothGrowth data in the R datasets package, perform some basic exploratory data analyses and provide a basic summary of the data. The approach of analysing the data will be to use confidence intervals and hypothesis tests to compare tooth growth by supplement and dose, as well as own conclusions and assumptions.

1. Load the ToothGrowth Data

```
library(datasets)
data(ToothGrowth)
```

2. Basic summary of the 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
```

```
str(ToothGrowth)
```

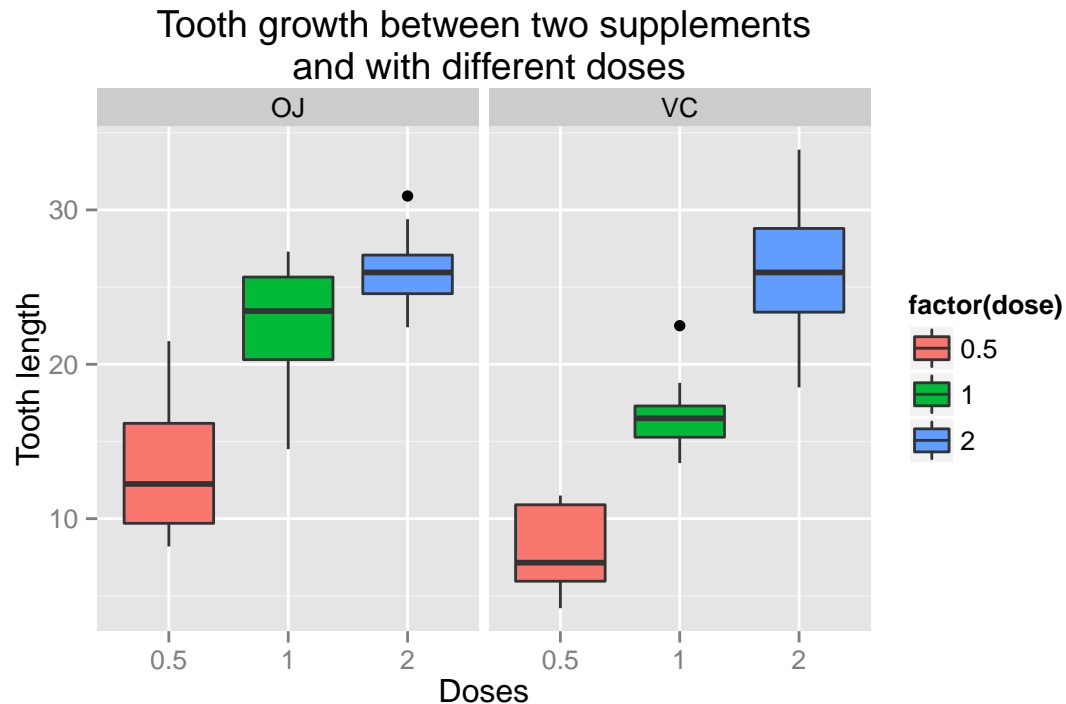
```
## 'data.frame':   60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

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

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

From the data we can see that we have 60 observations, split into two groups of supplementary (OJ and VC), which in turn are divided over three different dosages (0.5, 1, 2).

Let's make a box plot to compare the tooth lengths for the two supplements based on different dosing:



From first view of the box plot it seems that the tooth length is a bit longer with the supplement OJ when dosing is 0.5 and 1.0 compared to VC, whereas for 2.0 it doesn't seem to be any significant difference. However, let's test this before making any conclusions.

3. t test and confidence interval

As we are working with a small set of data we will use the Gosset's t distribution to test if the differences are significant between the supplement types and dosing for tooth growth:

$$H_0: \mu_{OJ|0.5} = \mu_{VC|0.5}$$

$$H_0: \mu_{OJ|1.0} = \mu_{VC|1.0}$$

$$H_0: \mu_{OJ|2.0} = \mu_{VC|2.0}$$

Subset the data based on dosage and run the t-test on the supplements:

```
#Dosing 0.5
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, subset(ToothGrowth, dose == 0.5))

##
## Two Sample t-test
##
## data: len by supp
## t = 3.1697, df = 18, p-value = 0.005304
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.770262 8.729738
## sample estimates:
## mean in group OJ mean in group VC
##          13.23          7.98
```

```
#Dosing 1.0
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, subset(ToothGrowth, dose == 1.0))
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = 4.0328, df = 18, p-value = 0.0007807
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 2.840692 9.019308
## sample estimates:
## mean in group OJ mean in group VC
## 22.70 16.77
```

```
#Dosing 2.0
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, subset(ToothGrowth, dose == 2.0))
```

```
##
## Two Sample t-test
##
## data: len by supp
## t = -0.0461, df = 18, p-value = 0.9637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.722999 3.562999
## sample estimates:
## mean in group OJ mean in group VC
## 26.06 26.14
```

Based on the `t.test` for the different doses, the confidence interval for the doses 0.5 and 1.0 exclude zero. Therefore, for the doses of 0.5 and 1.0, the null hypotheses are rejected, as the difference between the supplements is significant.

For the dose 2.0, the confidence interval include zero. Therefore, for the dose of 2.0, we fail to reject the null hypothesis.

4. Conclusion and assumptions.

There is a significant difference in the supplement types for doses of 0.5 and 1.0 on tooth growth.

There is no significant difference in the supplement types for doses of 2.0 on tooth growth.

We assume that the samples are:

- independent and identically distributed (iid),
- unpaired,
- equal variance &
- no other influencing factors to the tooth growth.