Inferential data analysis

Morgane Flauder

December 27, 2015

Analysis of the ToothGrowth data

In this assignment, we are going to analyze the ToothGrowth data in the R datasets package. This dataset show the effect of Vitamin C on tooth growth in Guinea Pigs.

Exploratory analysis

Let's look at the data.

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

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
```

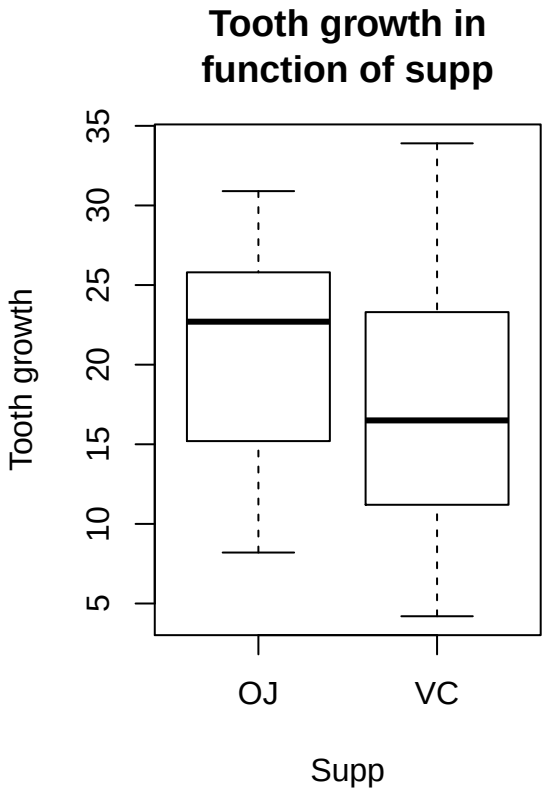
We can see that there is 3 variables : len, the variable of interest (the length of the tooth), supp (the supplement type : orange juice (OJ) or ascorbic acid (VC)), and the dose in mg.

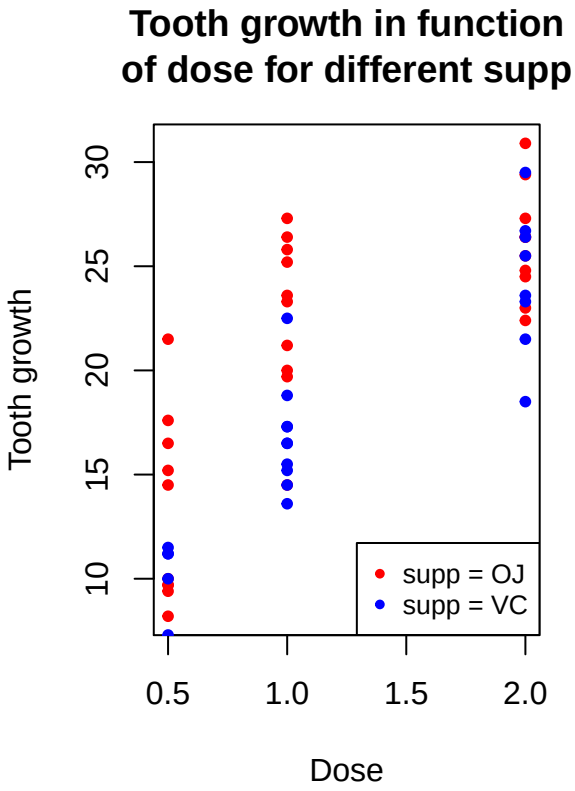
```
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

We can see that for each type of supplement, 3 doses (0.5, 1 and 2mg) were tested on 10 subject each.

Let's plot the length of the tooth in function of the supplement type and the dose.





We can see that on average, orange juice (OJ) supplement seems to induce more tooth growth, but for ascorbic acid (VC), there are more extreme values (minimum and maximum). We can also see that higher doses seem to be more effective.

Let's check theses theories with statistical methods.

Comparison of tooth growth by supp

We assume that the variance of the length is not equal for the two supplement types. Indeed, there are more extreme values for the ascorbic acid (VC). We also assume that the data aren't paired : the subjects are entirely different. Since the sample size is small (30 subject for each supplement type), we are going to perform a Student Test.

```
t.test(len ~ supp,
       paired = FALSE,
       var.equal = FALSE,
       data = ToothGrowth)$p.value
```

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

The p-value is superior to 0.05, therefore we can't conclude that the means are significantly different with 95% confidence. We conclude that there is no difference between the two supplement types.

Comparison of tooth growth by dose

We are now going to test the difference between the dose of Vitamin C. We are going to compare two doseS : 0.5 and 2mg. Let's assume that the variances are equal and the data aren't paired. There are 20 subjects for each dose, so we are going to perform a Student Test.

```
t.test(len ~ dose,
       paired = FALSE,
       var.equal = TRUE,
       data = ToothGrowth[ToothGrowth$dose != 1,])$p.value
```

```
## [1] 2.837553e-14
```

```
t.test(len ~ dose,
       paired = FALSE,
       var.equal = TRUE,
       data = ToothGrowth[ToothGrowth$dose != 1,])$conf[1:2]
```

```
## [1] -18.15352 -12.83648
```

We can see that the p-value is very small and the confidence interval is negative and between -18.2 and -12.8. There is a significant difference between the two means with 95% confidence. We can conclude that the 0.5mg dose is significantly less effective than the 2mg dose.

Comparison of tooth growth by dose for a specific dose

We are now going to compare the tooth growth for the different supplement types and for the 1mg dose. We assume that the variances are equal and the data aren't paired.

```
t.test(len ~ supp,
       paired = FALSE,
       var.equal = TRUE,
       data = ToothGrowth[ToothGrowth$dose == 1,])$p.value
```

```
## [1] 0.0007807262
```

```
t.test(len ~ supp,
       paired = FALSE,
       var.equal = TRUE,
       data = ToothGrowth[ToothGrowth$dose == 1,])$conf[1:2]
```

```
## [1] 2.840692 9.019308
```

The p-value is inferior to 0.05 and the confidence interval is positive and between 2.8 and 9.0, therefore there is a significant difference between the two mean with 95% confidence. We can conclude that with a 1mg dose, orange juice (OJ) is more effective than ascorbic acid (VC).