

# Statistical Inference Project - Part II

Analysis of the ToothGrowth data

*Loïc BERTHOU*

*August 19, 2015*

## Overview

This document is a basic analysis of the ToothGroth dataset provided in the R datasets packages.

## Introduction

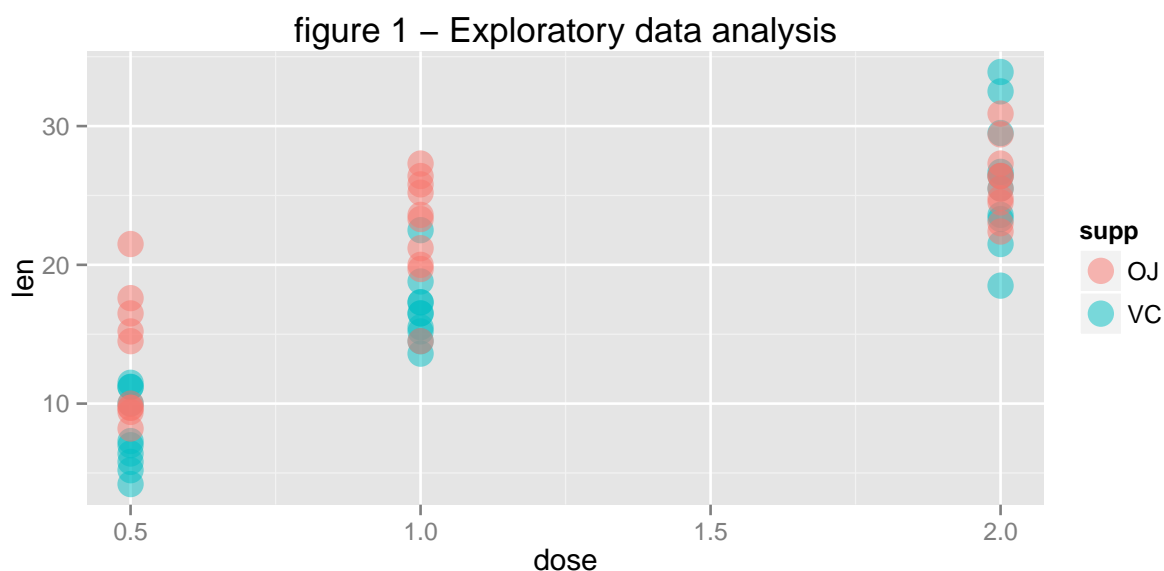
Since we don't have any knowledge about this dataset, we will first perform an exploratory analysis to get a better sense of the type of data we will be dealing with. Then we will perform a comparison of the tooth groth by supp and dose as requested.

## Exploratory data analysis

The first thing to do with the dataset is to read the documentation attached to it (throught the *help* function). The *str* function also provides a useful overview of the dataset:

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

Finally, I will draw a first plot of the scattered points (Figure 1).



## Summary of the data

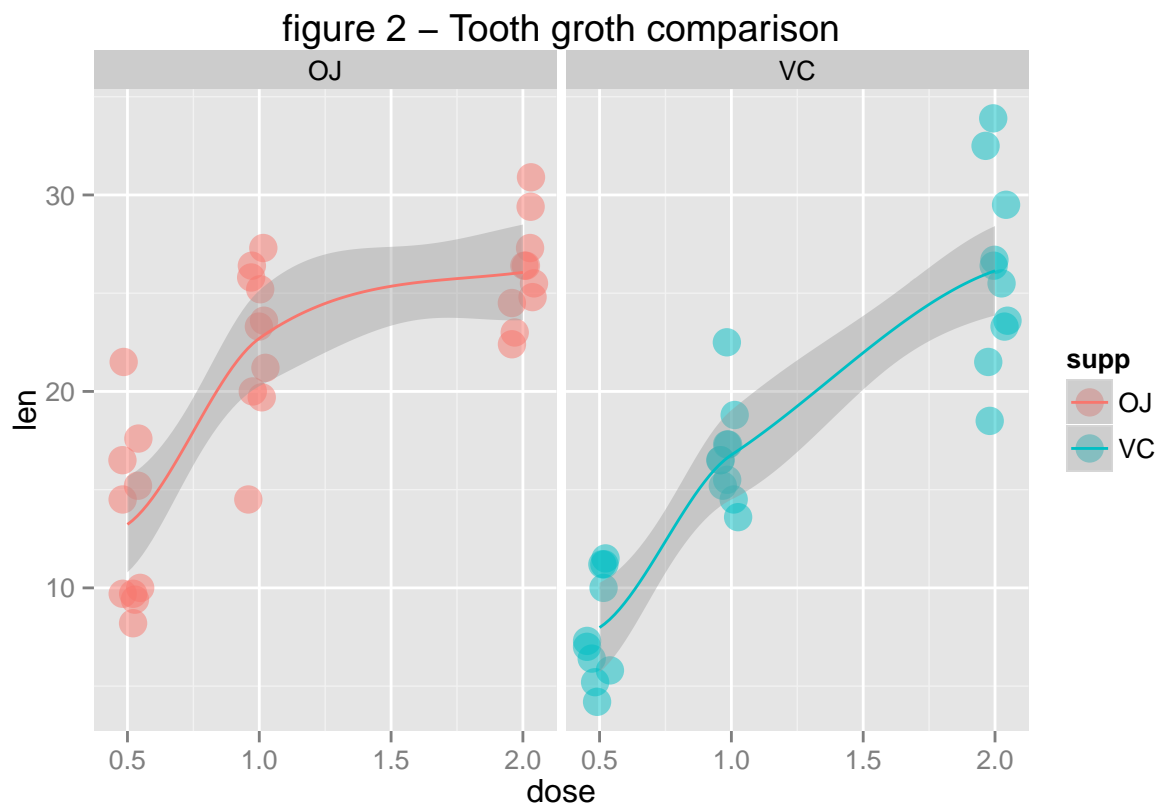
We have now a first idea of the data manipulated. These numbers have been produced by a study on the effect of Vitamin C on guinea pigs and more precisely the length of their teeth. The vitamin C has been administered either in the form of orange juice (represented "OJ") or ascorbic acid (represented "VC"). Each form has been administered at 3 dose levels to 10 animals (60 different animals in total).

## Tooth growth comparison by supp and dose

We will compare the effects of the orange juice and ascorbic acid for each dose. In this experience, each measure has been taken on a different animal, so we can not infer a grouping by animal.

We will then compare the mean length progression between two doses and look at the confidence interval to make any conclusion.

To illustrate the analysis, we draw the scattered points with the average curve.



### Orange Juice

#### From dose 0.5 to 1.0

The average growth difference is 9.47 with a confidence interval of 95% between 5.52 and 13.42.

### **From dose 1.0 to 2.0**

The average growth difference is 3.36 with a confidence interval of 95% between 0.19 and 6.53.

We can say that there is a difference between the effects of the two doses but are not guaranteed to be very big.

### **From dose 0.5 to 2.0**

The average growth difference is 12.83 with a confidence interval of 95% between 9.32 and 16.34.

The confidence interval with the confidence interval of the measurements “From dose 0.5 to 1.0” overlap. See the Conclusion.

## **Ascorbic Acid**

### **From dose 0.5 to 1.0**

The average growth difference is 8.79 with a confidence interval of 95% between 6.31 and 11.27.

### **From dose 1.0 to 2.0**

The average growth difference is 9.37 with a confidence interval of 95% between 5.69 and 13.05.

### **From dose 0.5 to 2.0**

The average growth difference is 18.16 with a confidence interval of 95% between 14.42 and 21.90.

## **Conclusion**

The effect of ascorbic acid seems pretty consistent. The effect of the intermediate dose (1.0 mg), which has a difference in growth with the weakest dose (0.5 mg) of 9.47. The strongest dose (2.0 mg) has an average difference in growth with the weakest dose (0.5 mg) of 18.16. By looking at the confidence intervals, we can say that the strongest dose has a very good chance of yielding better growth.

On the other hand, the effect of orange juice is not as important with a stronger dose. The effect of the intermediate dose (1.0 mg), which has a difference in growth with the weakest dose (0.5 mg) of 9.47. The strongest dose (2.0 mg) has an average difference in growth with the weakest dose (0.5 mg) of 12.83. By looking at the confidence intervals, we can say that the stronger dose has an effect, but it is not guaranteed to be much more effective than the intermediate dose.

Something interesting to note is that, as the average differences in growth are not behaving similarly, we can see that the orange juice is more efficient at smaller doses than ascorbic acid. So despite a steeper curve for the ascorbic acid, the absolute effects on the guinea pigs are very similar at 2.0 mg. This is clearly visible in *figure 2*.

## Appendix

### Code for generating figure 1

```
g <- ggplot(data=ToothGrowth, aes(x=dose, y=len))
g <- g + geom_point(aes(color=supp), size=5, alpha=.5)
g <- g + ggtitle("figure 1 - Exploratory data analysis")
g
```

### Code for generating figure 2

```
my_ci <- function(x) data.frame(y=mean(x), ymin=mean(x) - sd(x), ymax=mean(x) + sd(x))

g <- ggplot(data=ToothGrowth, aes(x=dose, y=len))
g <- g + geom_point(aes(color=supp), size=5, alpha=.5, position = position_jitter(w = 0.05, h = 0))
g <- g + geom_smooth(aes(color=supp))
g <- g + facet_wrap(~supp)
g <- g + ggtitle("figure 2 - Tooth groth comparison")
g
```

### Comments on figure 2

The points have been jittered horizontally to be able to see their concentration more clearly.