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 check if there is any sort of documentation attached to it.

```
help("ToothGrowth")
```

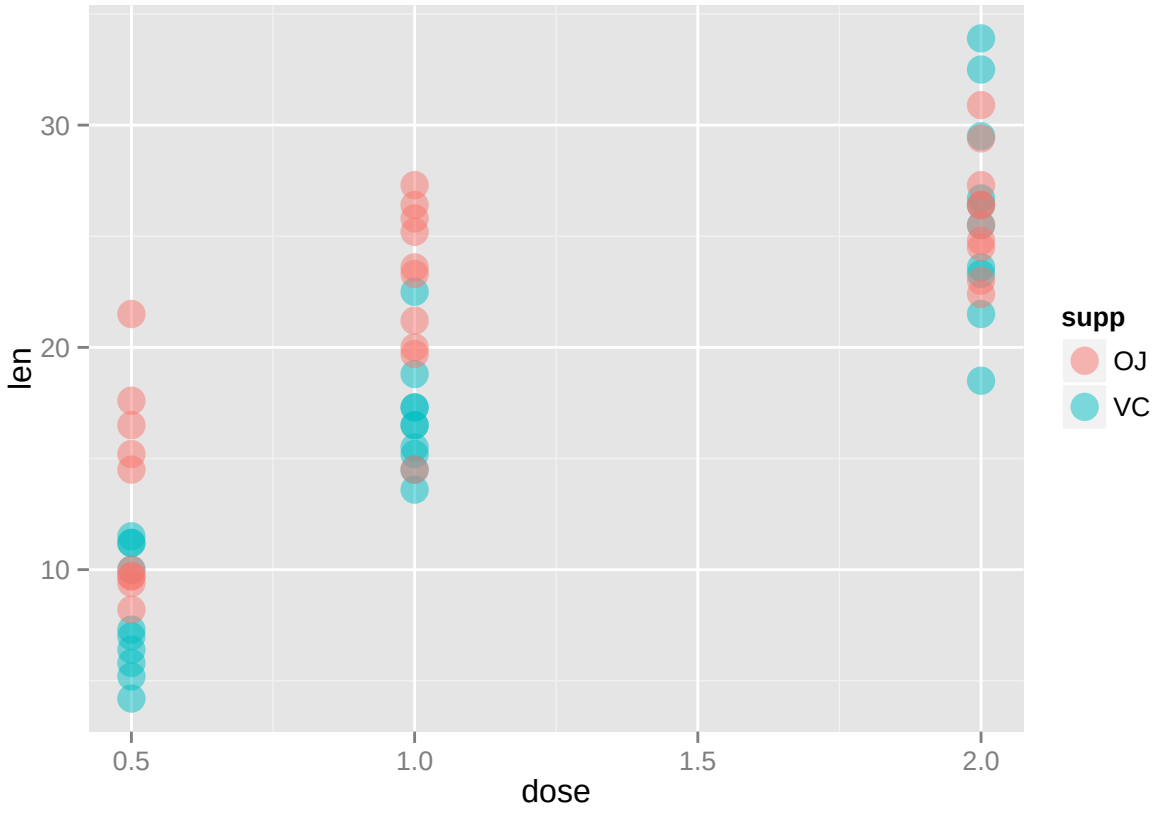
The *str* function returns some useful information.

```
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 ...
```

Finally, I will draw a first plot of the scattered points.

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



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

TODO

Conclusion

TODO