

ToothGrowth Exploratory Analysis

Axel Garcia

21 de junio de 2015

Overview

This document discusses the analysis done on the ToothGrowth data provided in R. That data set was obtained from C. I. Bliss (1952) The Statistics of Bioassay. Academic Press (according with the R Documentation).

That data set is mean to describe the effect of vitamin C on tooth growth in ginea pigs. More information about this data can be found in the R Documentation. The Following link points to it: [The Effect of Vitamin C on Tooth Growth in Guinea Pigs](#)

Data Information

Format

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length

[,2] supp factor Supplement type (VC or OJ).

[,3] dose numeric Dose in milligrams.

```
data(ToothGrowth)
```

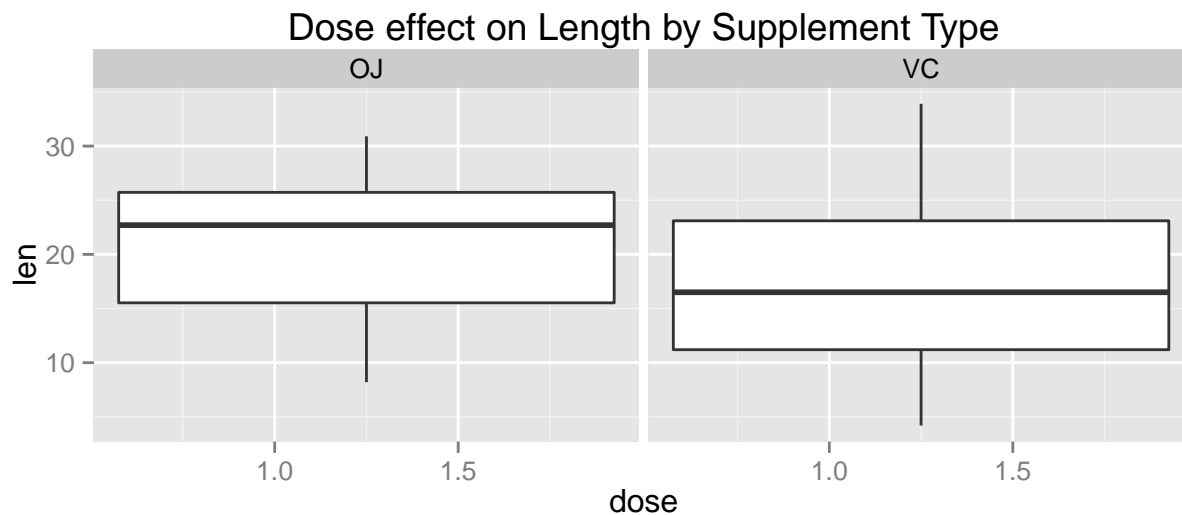
```
str(ToothGrowth)
```

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

Supp is a factor variable whereas the other two are numeric variables; furthermore, those numeric variables are continuous.

Wondering about the data, it maybe that the dose has something to do with the teeth length. The following plot shows how the relation ship between those two variables looks like:

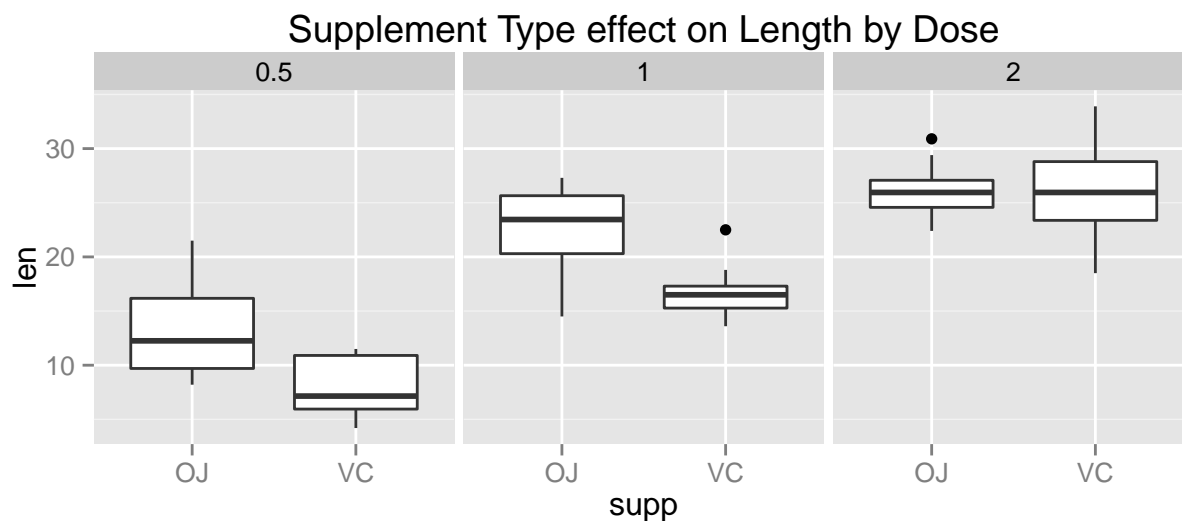
```
ggplot( ToothGrowth, aes(y=len, x=dose) ) +
  geom_boxplot() + facet_grid( .~supp ) +
  ggtitle( "Dose effect on Length by Supplement Type" )
```



The mean of both variables is contained by the 2nd and 3rd quartiles. That means that both data overlaps. Given that evidence, the dose does not seem to have a significant effect on the teeth length.

In the other hand, the supplement type may have an impact on the teeth length. Let's see how the relationship between those two variables looks like.

```
ggplot( ToothGrowth, aes(y=len, x=supp) ) +
  geom_boxplot() + facet_grid( .~dose ) +
  ggtitle( "Supplement Type effect on Length by Dose" )
```



When the dose is 0.5 the chart apparently shows that the means of OJ and VC supplement types are not contained in each other. However, there is still an overlap between the 1st quartile of OJ and the 3rd quartile of VC. Let's see what additional information the hypothesis test provides for this case.

When dose is 2, the plot shows that OJ data is completely within the VC distribution. In this case it does not seem to be any difference between them.

Finally, with dose equals to 1 the chart shows a clear difference and suggests OJ may be related to a bigger teeth growth than VC.

Hypothesis Test

Dose effect on Length by Supplement Type

The T-Test shows that there is not a significant statistical difference when comparing the effect that the supplement type had over tooth length:

```
test <- t.test( len~supp, data=ToothGrowth )
test

##
## Welch Two Sample t-test
##
## data: len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1710156 7.5710156
## sample estimates:
## mean in group OJ mean in group VC
## 20.66333 16.96333
```

Confidence interval includes 0 which means it is possible that it is equal. Also the p-value is 6.1% which is greater than 5%. The null hypothesis cannot be rejected.

Supplement Type effect on Length by Dose

```
data <- subset( ToothGrowth, dose %in% c(0.5) )
test <- t.test( len~supp, data=data )
```

Dose Equals to 0.5 P-Value of the test is 0.006 which is lower than 5%. for a dose equals to 0.5mm there is a significant statistic difference.

Dose Equals to 1 There seems to be an interesting relationship between the supplement type when the dose is one.

```
data <- subset( ToothGrowth, dose %in% c(1) )
test <- t.test( len~supp, data=data )
```

P-Value of the test is 0.001 this is lower than 5%, too. There is also a significant statistic difference when the dose equals to 1mm .

```
data <- subset( ToothGrowth, dose %in% c(0.5, 1) )
t.test( len~supp, data=data )
```

Dose Between 0.5 and 1

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##           17.965           12.375
```

Since p-value of the test is lower than 5% the t-test supports the alternative hypothesis. The effects of VC and OJ supplement types are different.

Conclusion

There is a significant statistical difference in the Guinea Pig teeth growth between VC and OJ supplement types when the dose is 0.5mg or 1mg. The second plot clearly shows that OJ supplement type has better results than VC supplement type.