Brief statistical data analysis on tooth growth data

Andra T.

31st of January 2016

Summary

In this project we aim to do a basic statistical analysis on the ToothGrowth data from the *datasets* package.

Data

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

The data is summarized in a data frame with 60 observations on 3 variables:

- len - tooth length
- supp - supplement type (VC - ascorbic acid or OJ -orange juice)
- dose - dose in milligrams per day

```r
library(datasets)
tooth<-ToothGrowth
head(tooth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```r
str(tooth)
```

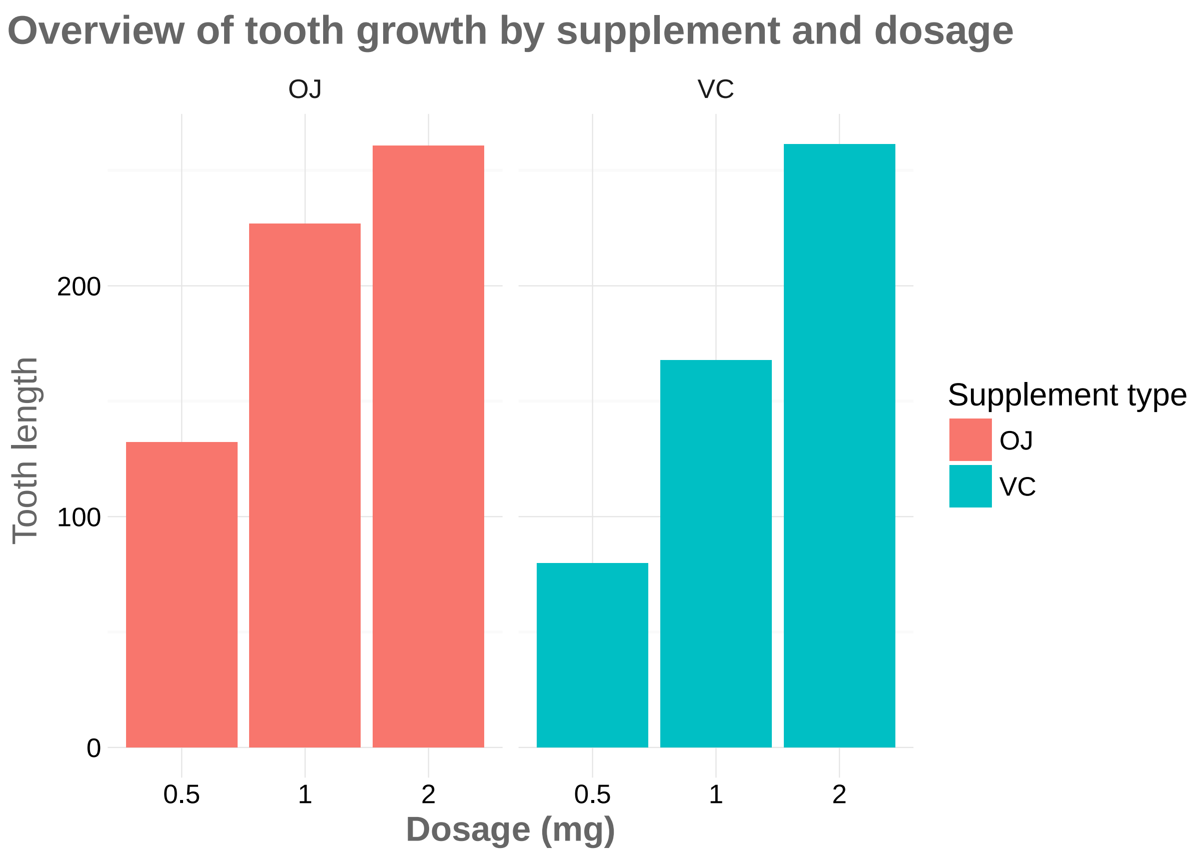
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Exploratory data analysis

The first step is to compare the tooth growth based on dosage between the two supplements .

```r
library(ggplot2)
library(dplyr)
```

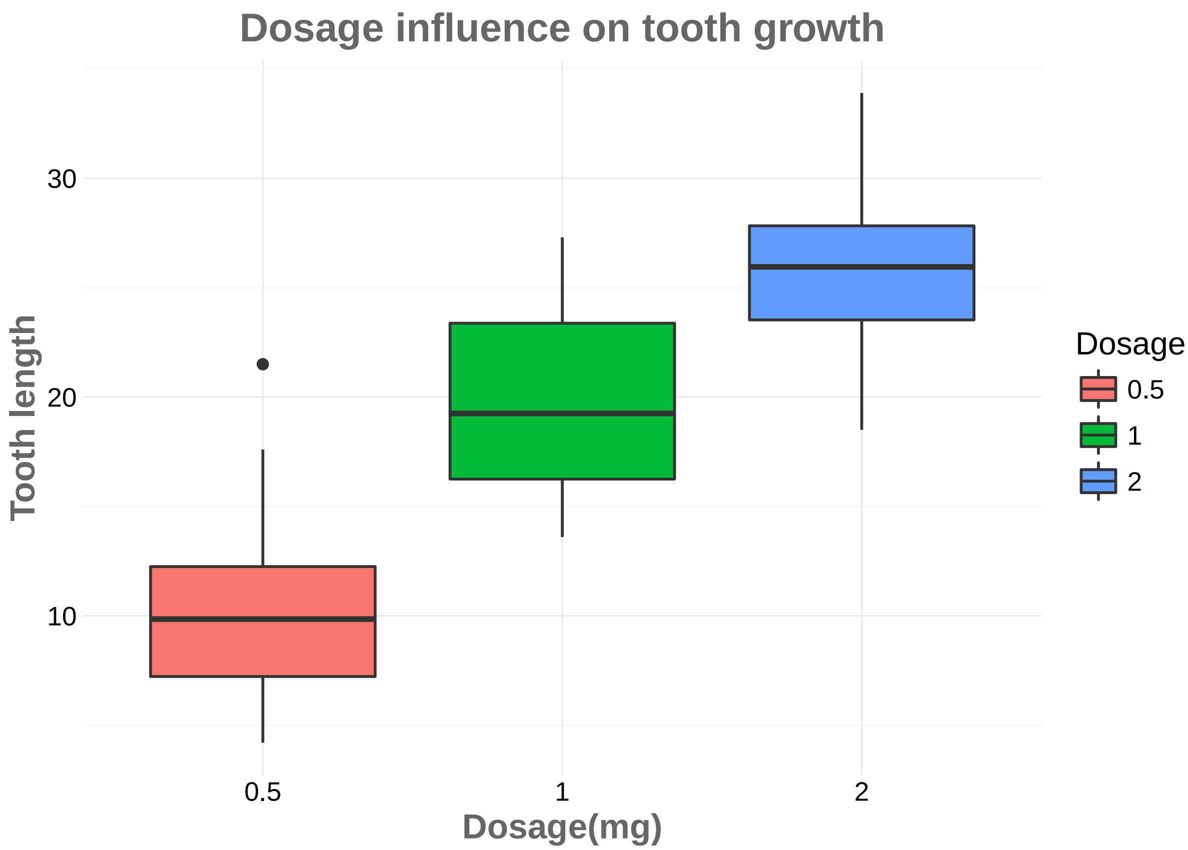
```
tooth$dose<-as.factor(tooth$dose)
ggplot(data=tooth, aes(x=dose, y=len, fill=supp)) + geom_bar(stat="identity",) +
facet_grid(. ~ supp)+
labs(x="Dosage (mg)", y=expression("Tooth length ")) + theme_minimal()+
ggtitle("Overview of tooth growth by supplement and dosage")+
theme(plot.title = element_text( color="#666666", face="bold", size=15, hjust=0.5)) +
theme(axis.title = element_text( color="#666666", face="bold", size=13)) +
guides(fill=guide_legend(title="Supplement type",face="bold", size=13))
```



We can observe from the graph above that overall, the guinea pigs that received Vitamin C from Orange juive have a higher level of tooth growth, regardless of the dosage - compared to the guinea pigs that received it from Ascorbic Acid.

Next, we would like to find out how the dosage alone is influencing the tooth growth, regardless of the mode of administration.

```
ggplot(aes(x=dose, y=len), data=tooth) + geom_boxplot(aes(fill=dose)) +
xlab("Dosage(mg)") +
ylab("Tooth length") + guides(fill=guide_legend(title="Dosage"))+theme_minimal()+
ggtitle("Dosage influence on tooth growth")+
theme(plot.title = element_text( color="#666666", face="bold", size=15, hjust=0.5)) +
theme(axis.title = element_text( color="#666666", face="bold", size=13))
```




```
tooth %>% group_by(dose) %>%
  summarize(
    mean= mean(len),
    q25 = quantile(len, 0.25),
    q75 = quantile(len, 0.75)
    ) %>%
  as.data.frame
```

```
##   dose   mean    q25    q75
## 1  0.5 10.605  7.225 12.250
## 2    1 19.735 16.250 23.375
## 3    2 26.100 23.525 27.825
```

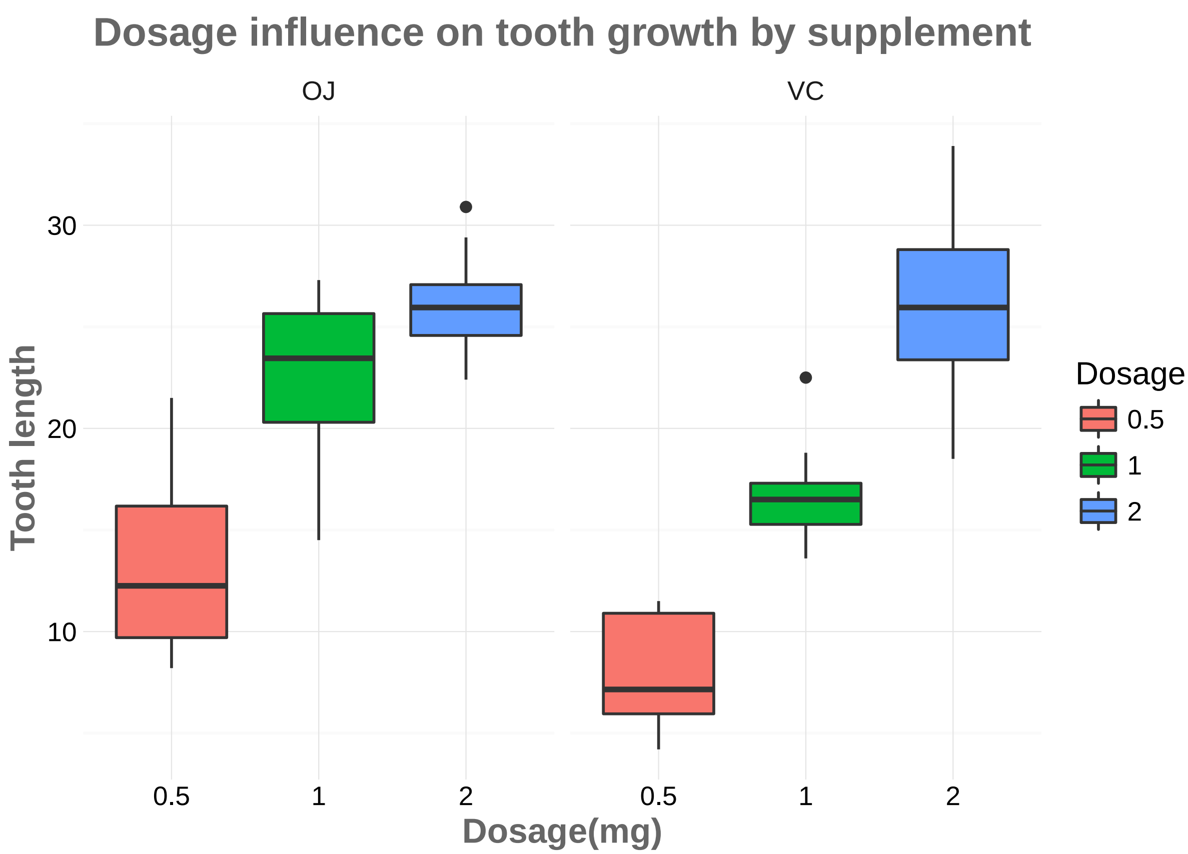
Although we have some outliers for the lowest dosage at tooth length ~20, we can observe that the highest the administered dose is, the highest the tooth growth.

However, there are very large differences between the three dosage groups (non overlapping interquartile ranges)

If we look at the dosage groups by supplement types, we can observe a similar increasing trend : dosage increases, tooth length increases regardless of supplement.

```
ggplot(aes(x=dose, y=len,fill=supp), data=tooth) + geom_boxplot(aes(fill=dose)) +
xlab("Dosage(mg)") +
facet_grid(. ~ supp)+
ylab("Tooth length") + guides(fill=guide_legend(title="Dosage"))+theme_minimal()+
```

```
ggtitle("Dosage influence on tooth growth by supplement")+
theme(plot.title = element_text( color="#666666", face="bold", size=15, hjust=0.5)) +
theme(axis.title = element_text( color="#666666", face="bold", size=13))
```



Hypothesis testing

Assumptions:

- Independence : the guinea pigs have been assigned randomly to the two groups (OJ and VC)
- Normal population from which the guinea pigs have been selected ; the population is representative for an entire population of guinea pigs.
- For hypothesis testing we assume that the variances are different for the two groups.

Supplement types affect tooth growth

HO: The difference in tooth growth means between OJ guinea pigs and VC guinea pigs is !=0

```
t.test(len ~ supp, ToothGrowth, var.equal = FALSE)
```

```
##
##  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
```

The p-value in this case is 0.06 which is greater than the 0.05 (significance threshold). The 95 percent confidence interval is [-0.1710156 , 7.5710156] , containing 0.

The null hypothesis cannot be rejected :

Conclusion 1 : Vitamin C supplement types do not affect tooth growth in guinea pigs.

Increased Vitamin C dosage levels are influencing tooth growth

In order to test the hypothesis above, we will look at the influence of increasing doses (in pairs) on tooth growth .

Case 1 : Dosages 0.5mg -0.1 mg

HO: The difference in tooth growth means between the guinea pigs with 0.5 mg and 1 mg Vitamin C administration is 0

```r
tooth1<-subset(tooth, tooth$dose==0.5 | tooth$dose==1)
t.test(len ~ dose, data = tooth1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Case 2: Dosages 0.5mg - 2mg

HO: The difference in tooth growth means between the guinea pigs with 0.5 mg and 2 mg Vitamin C administration is 0

```r
tooth2<-subset(tooth, tooth$dose==0.5 | tooth$dose==2)
t.test(len ~ dose, data = tooth2)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Case 3: Dosages 1mg - 2mg

HO: The difference in tooth growth means between the guinea pigs with 1 mg and 2 mg Vitamin C administration is 0

```r
tooth3<-subset(tooth, tooth$dose==1 | tooth$dose==2)
t.test(len ~ dose, data = tooth3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The p-values in all three cases tested above are much lower than the significance level 0.05. In this case we reject the null hypotheses listed in each case.

Conclusion 2: Increasing the Vitamin C dosage level will positively affect tooth growth in guinea pigs.