

# Statistical Inference Course Project - Data Analysis

*Daniel Pont*

*2018-09-30*

## Synopsis

This report is the second part of the Statistical Inference Course Project. It's a basic inferential data analysis of the ToothGrowth dataset that shows that vitamin C increases tooth growth in guinea pigs. It also shows that for dose < 2 tooth growth is higher with OJ than with VC.

## Basic Inferential Data Analysis of the dataset ToothGrowth

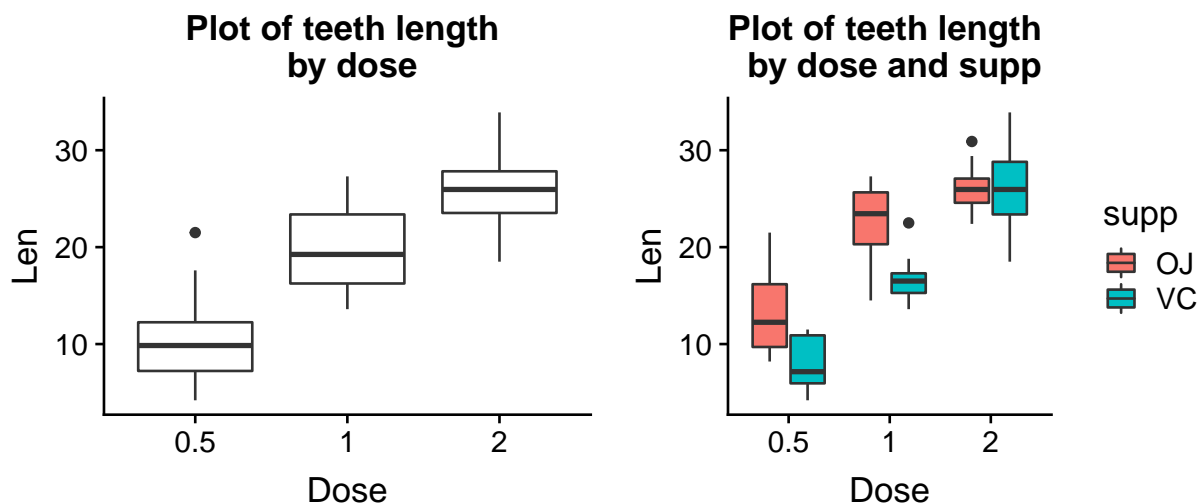
### 1) Exploratory Data Analysis

```
ToothGrowth$dose_factor <- as.factor(ToothGrowth$dose)

p_all <- ggplot(ToothGrowth, aes(x=dose_factor, y=len)) +
  geom_boxplot() +
  ggtitle("Plot of teeth length \n by dose") +
  xlab("Dose") + ylab("Len")

p_by_supp <- ggplot(ToothGrowth, aes(x=dose_factor, y=len, fill=supp)) + geom_boxplot() +
  ggtitle("Plot of teeth length \n by dose and supp") +
  xlab("Dose") + ylab("Len")

plot_grid(p_all, p_by_supp,
  ncol = 2, nrow = 1)
```



- Tooth growth seems to increase with vitamin C dose
- Tooth growth seems to increase more with OJ than with VC for dose < 2
- Tooth growth doesn't seem different with OJ or VC for dose = 2

## 2) Summary of the data

```
ToothGrowth %>% group_by(supp,dose) %>%  
  summarize(mean=mean(len), sd=sd(len))
```

```
## # A tibble: 6 x 4  
## # Groups:   supp [?]  
##   supp   dose mean    sd  
##   <fct> <dbl> <dbl> <dbl>  
## 1 OJ     0.5  13.2  4.46  
## 2 OJ     1    22.7  3.91  
## 3 OJ     2    26.1  2.66  
## 4 VC     0.5   7.98  2.75  
## 5 VC     1    16.8  2.52  
## 6 VC     2    26.1  4.80
```

## 3) Comparison of tooth growth by supp and dose

According to 2) “Summary of the data”, the variance of each group is different. We assume the groups are not paired. So we can perform t-tests with (paired=FALSE,var.equal=FALSE) for the following couple of datasets :

- len for dose = 2 vs len for dose = 0.5 (all supp)
- len for supp = OJ vs len for supp = VC (dose <2)
- len for supp = OJ vs len for supp = VC (dose =2)

```
dose0.5 <- ToothGrowth %>% filter(dose==0.5) %>% select(len)  
dose2.0 <- ToothGrowth %>% filter(dose==2) %>% select(len)
```

```
t.test(dose2.0,dose0.5,paired=FALSE)
```

```
##  
## Welch Two Sample t-test  
##  
## data: dose2.0 and dose0.5  
## 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:  
## 12.83383 18.15617  
## sample estimates:  
## mean of x mean of y  
## 26.100 10.605
```

```
OJ.low.dose <- ToothGrowth %>% filter(supp=="OJ",dose<2) %>% select(len)  
VC.low.dose <- ToothGrowth %>% filter(supp=="VC",dose<2) %>% select(len)
```

```
t.test(OJ.low.dose,VC.low.dose,paired=FALSE)
```

```
##  
## Welch Two Sample t-test  
##  
## data: OJ.low.dose and VC.low.dose  
## 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 of x mean of y
## 17.965 12.375

OJ.high.dose <- ToothGrowth %>% filter(supp=="OJ",dose==2) %>% select(len)
VC.high.dose <- ToothGrowth %>% filter(supp=="VC",dose==2) %>% select(len)

t.test(OJ.high.dose,VC.high.dose,paired=FALSE)

##
## Welch Two Sample t-test
##
## data: OJ.high.dose and VC.high.dose
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.79807 3.63807
## sample estimates:
## mean of x mean of y
## 26.06 26.14
```

To sum-up :

len means difference	95% Confidence Interval	P-value
dose=2 - dose=0.5	[12.8,18.2]	4.4e-14
OJ - VC (dose <2)	[1.88,9.30]	0.0042
OJ - VC (dose =2)	[-3.80,3.64]	0.96

#### 4) Conclusion

Under the following assumptions :

- the t-tests with default variance parameter (unequal variance) is relevant
- groups are not paired (different guinea pigs are used for different supp and dose combinations)

We found that :

- Tooth growth is significantly increased when dose is increased from 0.5 to 2 (p-value << 0.05 and the 95% confidence interval minimum is strictly positive)
- Tooth growth is significantly increased for dose <2 when OJ is used instead of VC (same reason than above)
- For dose =2, there's no significant difference in tooth growth between OJ and VC ( p-value >> 0.05, confidence interval centered on 0)