

## Analyses of tooth growth

Author: “G ARGENTON”

### Overview

Analyses of tooth growth as part of “Course project” of Statistical Inference course of John Hopkins university in Coursera

#### Exploratory data analysis

Let's load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
summary(ToothGrowth)
```

```
##      len      supp      dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07  VC:30   1st Qu.:0.500
##  Median :19.25
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

#### Summary of the data

Let's analyse the average tooth growth by sub groups of specific dose and supp  
Let's plot the data to see patterns

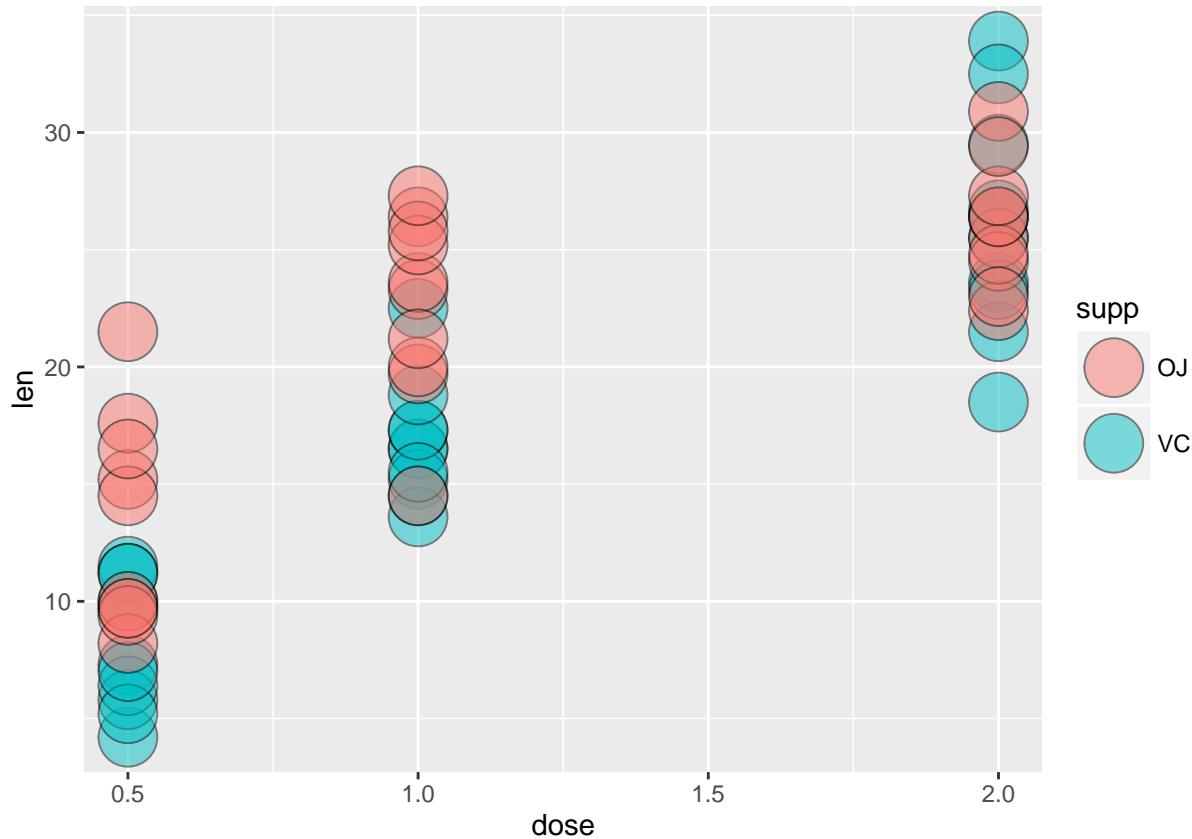
```
ToothGrowthByDoseSupp <- aggregate(len ~ dose + supp, data=ToothGrowth, mean)
ToothGrowthByDoseSupp
```

```
##   dose supp   len
## 1  0.5   OJ 13.23
## 2  1.0   OJ 22.70
## 3  2.0   OJ 26.06
## 4  0.5   VC  7.98
## 5  1.0   VC 16.77
## 6  2.0   VC 26.14
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
g <- ggplot(data = ToothGrowth, aes(x = dose, y = len, fill = supp ))
g <- g + geom_point(size =10, pch = 21, alpha = .5)
g
```



It seems that tooth growth is increasing with dosis, and that OJ is more efficient than VC.

#### Comparaison on tooth growth by supp and dose.

For supp VC, let's compare the group with 0.5 dosis and the group with 2 dosis

```
# subsetting data into two groups g1 & g2
g1 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5,]$len
g2 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2,]$len
print(paste("Average tooth growth supp VC dose 0.5 :", mean(g1), ", dose 2 :", mean(g2)))

## [1] "Average tooth growth supp VC dose 0.5 : 7.98 , dose 2 : 26.14"

print(paste("Average increase of tooth growth using VC increasing dosis from 0.5 to 2 :", mean(g2)-mean(g1)))

## [1] "Average increase of tooth growth using VC increasing dosis from 0.5 to 2 : 18.16"

# Let's calculate a 95 % student's t confidence interval for two independant groups
# we assume constant variance
sd1 <- sd(g1); sd2 <- sd(g2)
pv <- (9*sd1^2+9*sd2^2)/18
semd <- sqrt(pv)*sqrt(1/10+1/10)
round(mean(g2)-mean(g1) + c(-1,1)*qt(0.975,18)*semd,2)

## [1] 14.49 21.83
```

Let's test the hypothesis of more efficiency with supp OJ than VC

```
# subsetting data into two groups VC & OJ
VC <- ToothGrowth[ToothGrowth$supp == "VC" ,]$len
OJ <- ToothGrowth[ToothGrowth$supp == "OJ", ]$len
print(paste("Average tooth growth supp VC :", round(mean(VC),2),"supp OJ:", round(mean(OJ),2)))

## [1] "Average tooth growth supp VC : 16.96 supp OJ: 20.66"

print(paste("Average increase of tooth growth using OJ vs VC :",mean(OJ)-mean(VC)))

## [1] "Average increase of tooth growth using OJ vs VC : 3.7"

# Let's test the hypothesis of one supp being more efficient (two sided test) with alpha = 5%
# the two groups are independant, and we assume constant variance
t.test(OJ, VC, paired = FALSE, var.equal = TRUE)

## 
## Two Sample t-test
##
## data: OJ and VC
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064 7.5670064
## sample estimates:
## mean of x mean of y
## 20.66333 16.96333

qt(p = 0.975,df = 58)

## [1] 2.001717
```

### Conclusions and the assumptions needed for conclusions.

Tooth length is higher with VC and dose = 2 than with VC and dose = 0.5 : with 95% confidence, increase is between 14.49 and 21.83

We can not conclude that “one supp is more efficient than the other” with alpha = 5% : we fail to reject the null hypothesis (O within 95% confidence interval, or statistic  $1.9153 < 2.0017$  t quantile, or p-value = 0.06 > 0.05)

NB : We have assumed that distribution is gaussian or at least symmetric and mound shaped so that t tests and Student's t confidence interval applies, and data are reasonably iid sample of the population. We also assume that variance is constant within the groups.