Statistical Inference Project Exponential distribution vs CLT

Alfonso Gonzalez Damian

Exploring dataset

- Exploring the dataset with summary, dim and plot.
- We can see two clear groups in tooth growth by supp and three groups by dose

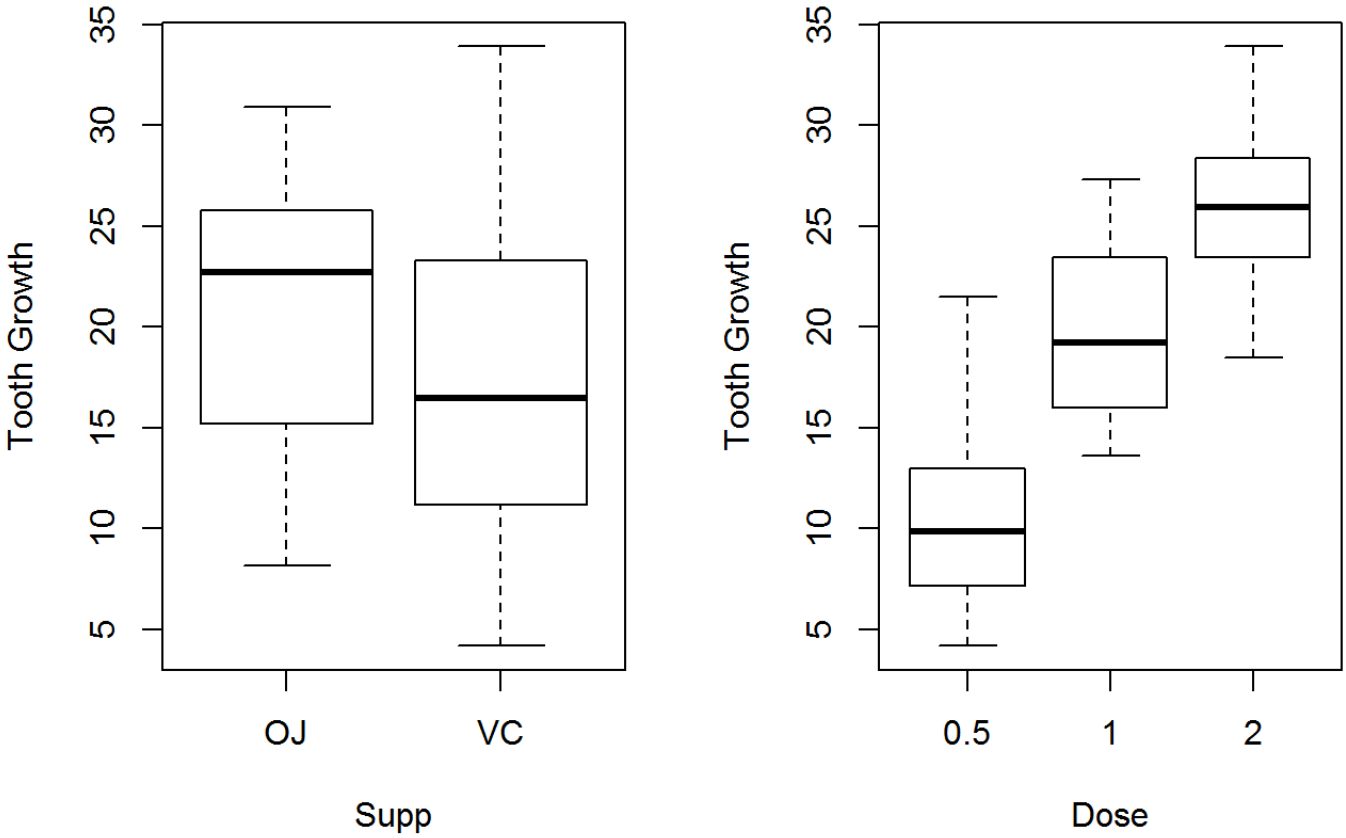
```
data(ToothGrowth)
library(reshape2)
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
n <- length(ToothGrowth$len)
par(mfrow =c(1,2))
plot(ToothGrowth$len ~ ToothGrowth$supp, type="l", xlab="Supp", ylab="Tooth Growth")
plot(ToothGrowth$len ~ as.factor(ToothGrowth$dose), xlab="Dose", ylab="Tooth Growth")
```



* We can test if this sample is a good estimator of the real population differences in toothgrowht

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

- Using the t.test in R for the hypothesis test that there is a difference greater than 0:

```
results <- t.test(len ~ supp, paired = FALSE, alternative="g", data= ToothGrowth)
results
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

- As we can see in the results the p-value < 0.05, so we can accept the alternative hypothesis

Conclusions and the assumptions needed

We can conclude that the confidence interval for the two groups (OC and VC) mean difference in toot growth with a 95% of confidence includes the true difference between them.