# Statiscal Inference Final Project

Maria Cristina Alameda Salas

25/4/2023

# Part 2: Basic Inferential Data Analysis

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

## Loading the ToothGrowth data

```
# load "tidyverse" plotting package
library(ggplot2)



# Load data
data(ToothGrowth)
```

## Performing some basic exploratory data analyses

In the following code, we can see that 3 columns compose our dataset: len, supp & dose.

```
# Show first rows
head(ToothGrowth)
```

```
##    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
```

```
dim(ToothGrowth)
```

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

## Provide a basic summary of the data.

This data contains 3 columns. The first column (*len*) talks about the length of tooth of guinea pigs. It is a continious value in range 4.20 - 33.90 with a mean of 18.81. The columns *supp* is about the supplement took by the animal. This column is discrete, with just two possible values (OJ & VC).

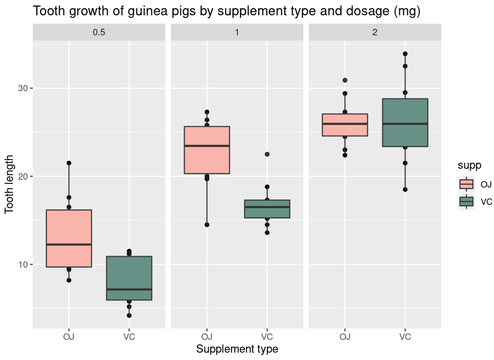
The last column (*dose*) shows data about dosage of supplement in mg. It goes from 0.5 mg to 2 mg.

```
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
```

Let's plot the previous information by *supp* group.

```
qplot(factor(supp), len, data = ToothGrowth, facets=~dose,
      main="Tooth growth of guinea pigs by supplement type and dosage (mg)",
      xlab="Supplement type", ylab="Tooth length") +
      geom_boxplot(aes(fill = supp)) +
      scale_fill_manual(values=c("#f9b4ab", "#679186"))
```




As it's plotted, total length of tooth increases as dosage is higher (2mg). In addition, there are difference between those animals given OJ supplement and VC one. In the specific case of VC, it starts from a lower value and achieve a great result after increasing dosage. OJ supplement increase tooth length as dosage increase as well. However, as starts from a higher value than VC, the increasing is minor.

## Hypothesis Test

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

### Assumptions

- Supposing the random variable of interest *Tooth length* has a know mean μ and a variance $\sigma^2$. We assume that *Tooth length* has a normal distribution, that is to say, X ~ N(μ, $\sigma^2$) State your conclusions and the assumptions needed for your conclusions.
- Variances of tooth growth are different when using different supplement and dosage.
- The three variables are independent and identically distributed (i.i.d.).

### Supplement Hypothesis

Null hypothesis: 'There is no difference in tooth growth when using supplements (OJ or VC) -> H0: lenOJ = lenVC

Alternate hypothesis: 'There are more tooth growth when using VC than OJ'

```
t.test(ToothGrowth$len[ToothGrowth$supp == 'OJ'], ToothGrowth$len[ToothGrowth$supp == 'VC'], alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

As the p-value (0.03032) is lower than 0.05, then we reject the null hypothesis.

### Supplement Hypothesis

Null hypothesis: 'There is no difference in tooth growth when using different dosage

Alternate hypothesis: 'There are more tooth growth when the dosage increases'

Comparing 0.5 to 1.0 mg

```
t.test(ToothGrowth$len[ToothGrowth$dose == 0.5], ToothGrowth$len[ToothGrowth$dose == 1], alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Comparing 1.0 to 2.0mg.

```
t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose == 2], alternative = "less", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 2]
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Both test show a p-values lower than 0.5. Then we reject the null hypothesis. This can be interpreted as, based on these low p-values, it is very likely that dosage has an effect on length, and a major value of dosage, major increase in length.

# Conclusions

We conclude that, due to the p-value obtained, there is a difference between administering the OJ supplement and the VC. In addition, there is a greater increase in tooth size when a higher dose of these supplements is administered.