# Course 6 - Project - Part 2

*Bruno Assis*

*19/03/2017*

## Impact of Vitamin C in the Tooth Growth of Guinea Pigs

### Overview

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

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

### Loading ToothGrowth Data and support libraries

```
library(ggplot2)
data("ToothGrowth")
```

### Exploratory Data Analyses

```
tooth.df <- data.frame(ToothGrowth)
dim(tooth.df)
```

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

```
head(tooth.df)
```

```
##    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
```

```
summary(tooth.df)
```

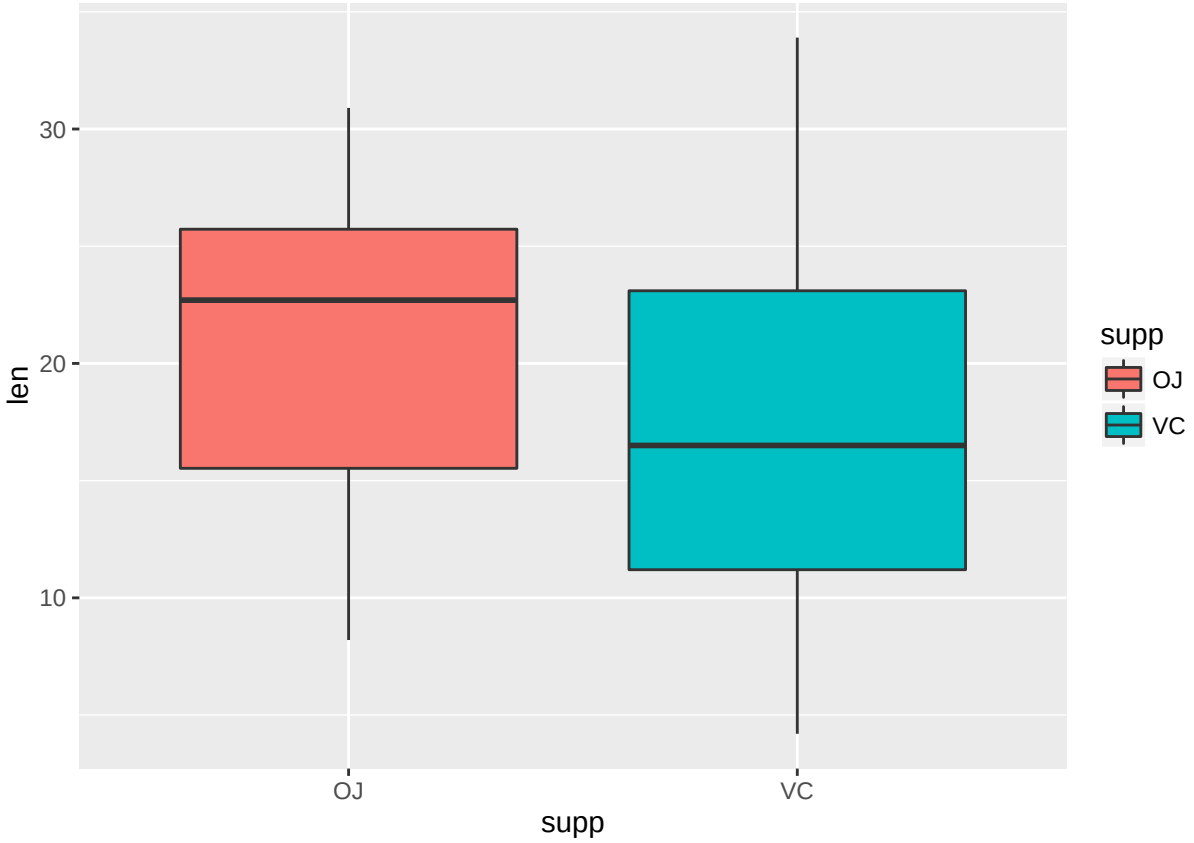
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Let's convert the `supplement` and `dose` attributes to factors:

```
tooth.df$supp <- factor(tooth.df$supp)
tooth.df$dose <- factor(tooth.df$dose)
```

Let's plot a boxplot comparing each `supplement` effect on the teeth's `length`:

```
ggplot(tooth.df, aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp))
```

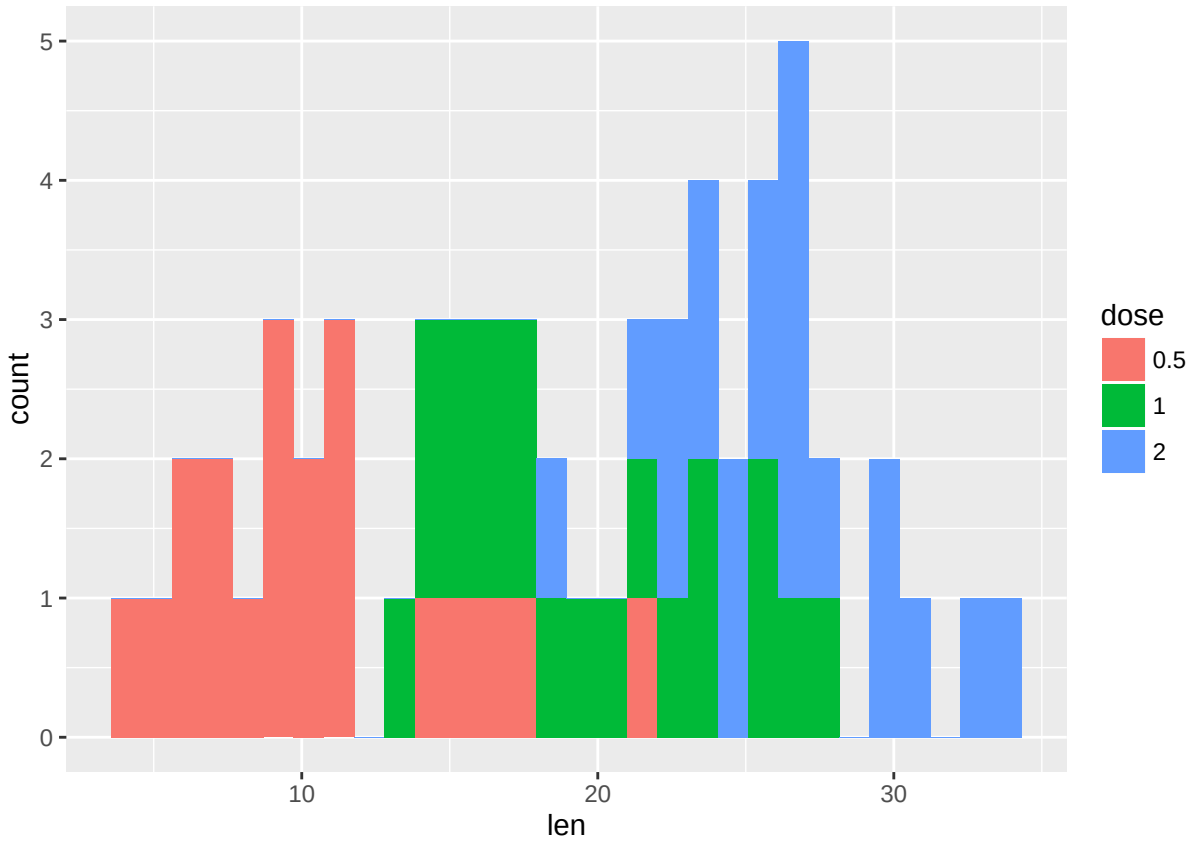



We see there's little difference between the `length` of each `supplement`.

Let's plot a histogram comparing each `dose` effect on the teeth's `length`:

```
ggplot(tooth.df, aes(x = len, fill = dose)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




We can see there is a relation between the given `dose` and the teeth's `length`.

### Comparing Supplements and their Doses to Teeth Length

Let's first compare the difference between supplements without looking at dosage:

```
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tooth.df)
```

```
##
##  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
```

Now let's compare the difference between dosages without looking at the supplement:

```
dosages <- unique(tooth.df$dose)
len.dosages <- length(dosages)

for (i in 1:(len.dosages-1)) {
    for (j in (i+1):len.dosages) {
      print(paste("T-Test between", dosages[i], "dosage and", dosages[j], "dosage"))
```

```
      print(t.test(subset(tooth.df, tooth.df$dose == dosages[i])$len,
                   subset(tooth.df, tooth.df$dose == dosages[j])$len,
                   paired = FALSE,
                   var.equal = FALSE))
    }
}
```

```
## [1] "T-Test between 0.5 dosage and 1 dosage"
##
##  Welch Two Sample t-test
##
## data:  subset(tooth.df, tooth.df$dose == dosages[i])$len and subset(tooth.df, tooth.df$dose == dosag
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
##
## [1] "T-Test between 0.5 dosage and 2 dosage"
##
##  Welch Two Sample t-test
##
## data:  subset(tooth.df, tooth.df$dose == dosages[i])$len and subset(tooth.df, tooth.df$dose == dosag
## 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 of x mean of y
##    10.605    26.100
##
## [1] "T-Test between 1 dosage and 2 dosage"
##
##  Welch Two Sample t-test
##
## data:  subset(tooth.df, tooth.df$dose == dosages[i])$len and subset(tooth.df, tooth.df$dose == dosag
## 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 of x mean of y
##    19.735    26.100
```

## Conclusions and Assumptions

**Assuming that the sample sizes are small and the t-test is appropriate**:

We can see there is no statistically significant difference between supplements, as their p.value is `0.06` and the confidence interval goes from `-0.17` to `7.57`, therefore, including `0`.

We also conclude there is statistically significant evidence that there is a correlation between dosage and

teeth length, by looking at the plot above and by analyzing the p.values of the tests, all of them very small (smaller than `0.00001`) and with confidence intervals that did not include `0`.