

Statistical Inference - Inferential Data Analysis

Silva, RAFAEL

December 4, 2017

Setup

This section presents all the libraries used during the project.

```
library(datasets)
library(ggplot2)
library(dplyr)
library(knitr)
```

Introduction

The objective of this project is to make some inferences about the length variation on each supplement and dose on the ToothGrowth data from the R datasets package. The analysis is going to follow the four steps below:

1. Exploratory data analysis and summary;
2. Hypothesis testing;
3. Conclusions and assumptions.

The chunk bellow is used to read the data set into the workspace.

```
data("ToothGrowth")
```

For more information on the data set (description, dimensions, variables and source) enter ?ToothGrowth on the console to check its documentation.

Exploratory analysis

To start the analysis, a brief verification on how the dataset looks like is necessary.

```
dim(ToothGrowth)
```

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

```
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
```

As the outputs above show, the data set has 60 observations on 3 variables. As the introduction chapter states, the objective is to make some inferences about the length variation on each supplement and dose. The boxplot below gives a feeling about how each combination affects the length.

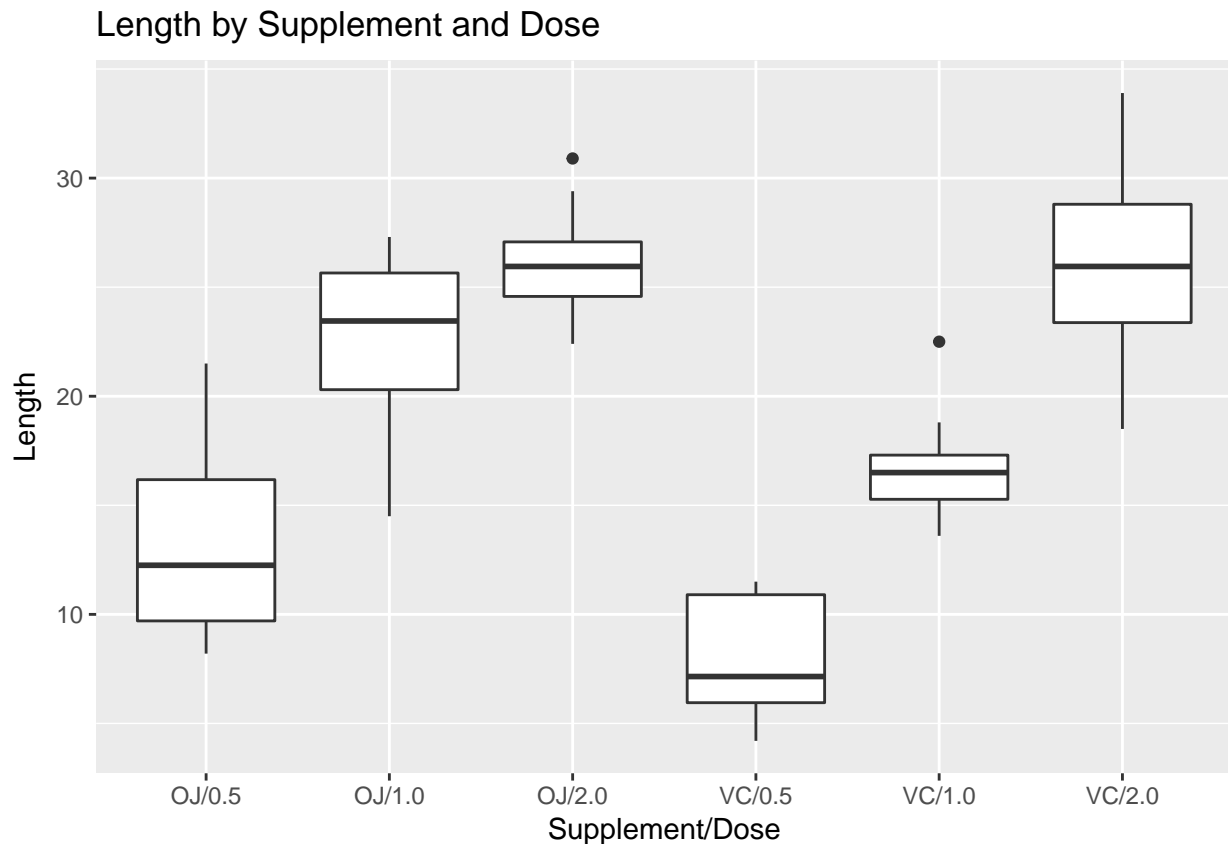
```

ToothGrowth <- ToothGrowth %>%
  mutate(suppDose = paste(supp, format(dose, digits = 2), sep = "/"))

g1 <- ggplot(ToothGrowth, aes(x = suppDose, y = len)) +
  geom_boxplot() +
  labs(title = "Length by Supplement and Dose",
       x = "Supplement/Dose", y = "Length")

print(g1)

```



It is important to notice that, in the graph above, the x axis does not present a continuous scale. This means that when analysing the graph you should not look at how the length increase “evolve” through the doses but at how each group behave in the y scale.

In the table below, the main statistical attributes of each group are summarized.

```

summ <- ToothGrowth %>%
  group_by(suppDose) %>%
  summarise(min = min(len),
            q1 = quantile(len, .25),
            mean = mean(len),
            median = median(len),
            q3 = quantile(len, .75),
            max = max(len),
            variance = var(len),
            n = length(len))

```

```
kable(summ, caption = "Data Summary by Supplement and Dose",
      col.names = c("", "Min", "1st Qu.", "Mean", "Median", "3rd Qu.",
                    "Max", "Variance", "n"))
```

Table 1: Data Summary by Supplement and Dose

	Min	1st Qu.	Mean	Median	3rd Qu.	Max	Variance	n
OJ/0.5	8.2	9.700	13.23	12.25	16.175	21.5	19.889000	10
OJ/1.0	14.5	20.300	22.70	23.45	25.650	27.3	15.295556	10
OJ/2.0	22.4	24.575	26.06	25.95	27.075	30.9	7.049333	10
VC/0.5	4.2	5.950	7.98	7.15	10.900	11.5	7.544000	10
VC/1.0	13.6	15.275	16.77	16.50	17.300	22.5	6.326778	10
VC/2.0	18.5	23.375	26.14	25.95	28.800	33.9	23.018222	10

Continue...