Statistical Inference Project Part 2

Paulo Cardoso

January 28, 2016

Overview

This is the second part of Statistical Inference class project that aims to perform a basic inferential analyses utilizing the ThoothGrowth data of the R datasets package.The dataset is compose by 60 observations, length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Set up

Loading the external libraries and the data.

```
# Loading libraries
library(datasets)
library(lattice)

# Loading the data
data(ToothGrowth)
```

Exploratory Data Analyses

In order to understand the data set is necessary that a few metrics be presented.

I)Presenting the struture of the dataset.

```
# Struture presentation
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

II)Display the amount of rows.

```
# Number of rows
nrow(ToothGrowth)
```

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

III)Display the first six rows.

```
# Head of the dataset
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
```

IV)Display the last six rows.

```
# Tail of the dataset
tail(ToothGrowth)
```

```
##     len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

V) Presenting the summary of the dataset.

```
# Summary
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
```

VI) Scatterplot showing every single value, and a boxplot chart summarizing the dataset.

```
# Scatterplot
xyplot(len~dose|supp, ToothGrowth,
       main="Scatterplots of type of supplement and dose",
       ylab="Length", xlab="Dose")
```

Scatterplots of type of supplement and dose
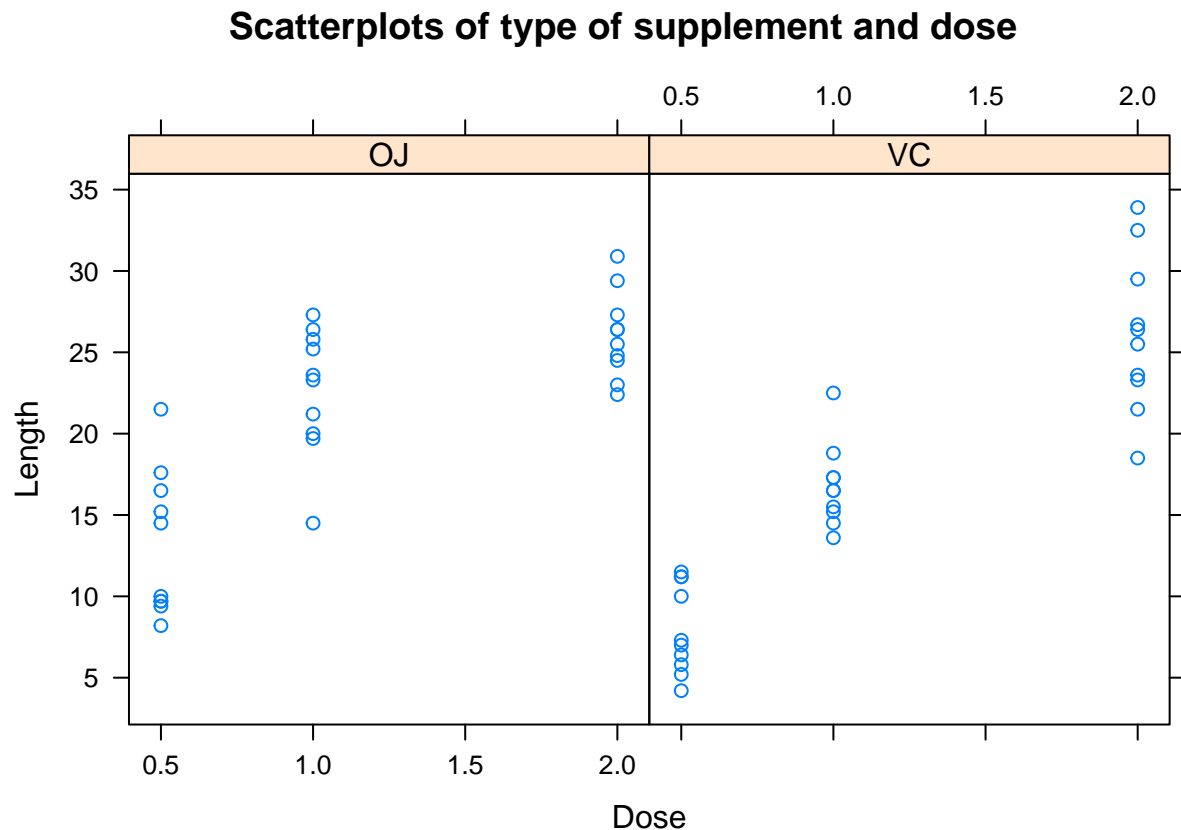
VII) Boxplot representing the tooth length by dose, on that way comparing the two delivery methods. And also a basic summary of each delivery method

```r
# Separating the two diferent delivery methods
oj <- levels(ToothGrowth$supp)[1]
vc <- levels(ToothGrowth$supp)[2]

# Boxplot
par(mfrow = c(1, 2))
boxplot(len ~ dose,
        ToothGrowth[ToothGrowth$supp == oj,],
        main="OJ",
        xlab="Dose",
        ylab="Tooth Length",
        ylim=range(0,35))
boxplot(len ~ dose,
        ToothGrowth[ToothGrowth$supp == vc,],
        main="VC",
        xlab="Dose",
        ylab="Tooth Length",
        ylim=range(0,35))
```

OJ

VC
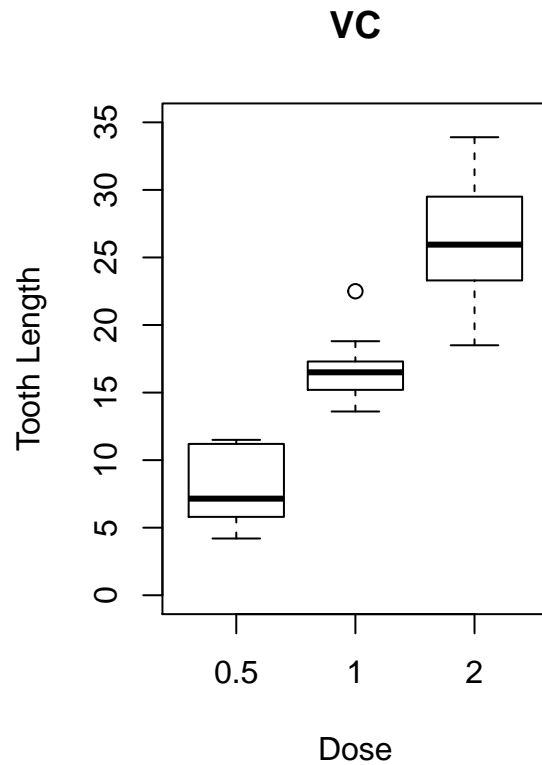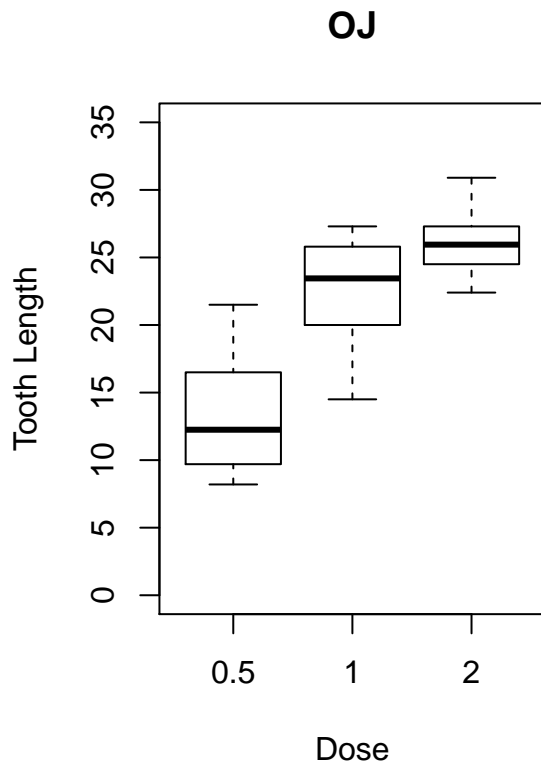

```r
# Summary from Orange Juice
summary(ToothGrowth[ToothGrowth$supp == oj,])
```

```
##       len          supp         dose
##  Min.   : 8.20   OJ:30   Min.   :0.500
##  1st Qu.:15.53   VC: 0   1st Qu.:0.500
##  Median :22.70           Median :1.000
##  Mean   :20.66           Mean   :1.167
##  3rd Qu.:25.73           3rd Qu.:2.000
##  Max.   :30.90           Max.   :2.000
```

```r
# Summary from Ascorbic Acid
summary(ToothGrowth[ToothGrowth$supp == vc,])
```

```
##       len          supp         dose
##  Min.   : 4.20   OJ: 0   Min.   :0.500
##  1st Qu.:11.20   VC:30   1st Qu.:0.500
##  Median :16.50           Median :1.000
##  Mean   :16.96           Mean   :1.167
##  3rd Qu.:23.10           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Inferential Statistics

Do the tooth length of the guinea pigs depends on delivery methods? To analyze tooth growth, we carry out two-sample (orange juice vs. ascorbic acid) T-tests at each dosage level.

The T-test at 0.5 mg yields the following:

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The T-test at 1 mg yields the following:

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 1, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The T-test at 2 mg yields the following:

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == 2, ])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Conclusion

From the T test performed above it can be concluded that from statistical significance p-value that the dosages of 0.5 mg and 1 mg of orange juice is more effective. Thus promoting greater dental growth that ascorbic acid. Analyzing the p-value of 2 mg can not be concluded that orange juice promotes dental growth more effectively than ascorbic acid.