Tooth Growth Analyses

JoÃ£o Pedro Schmitt

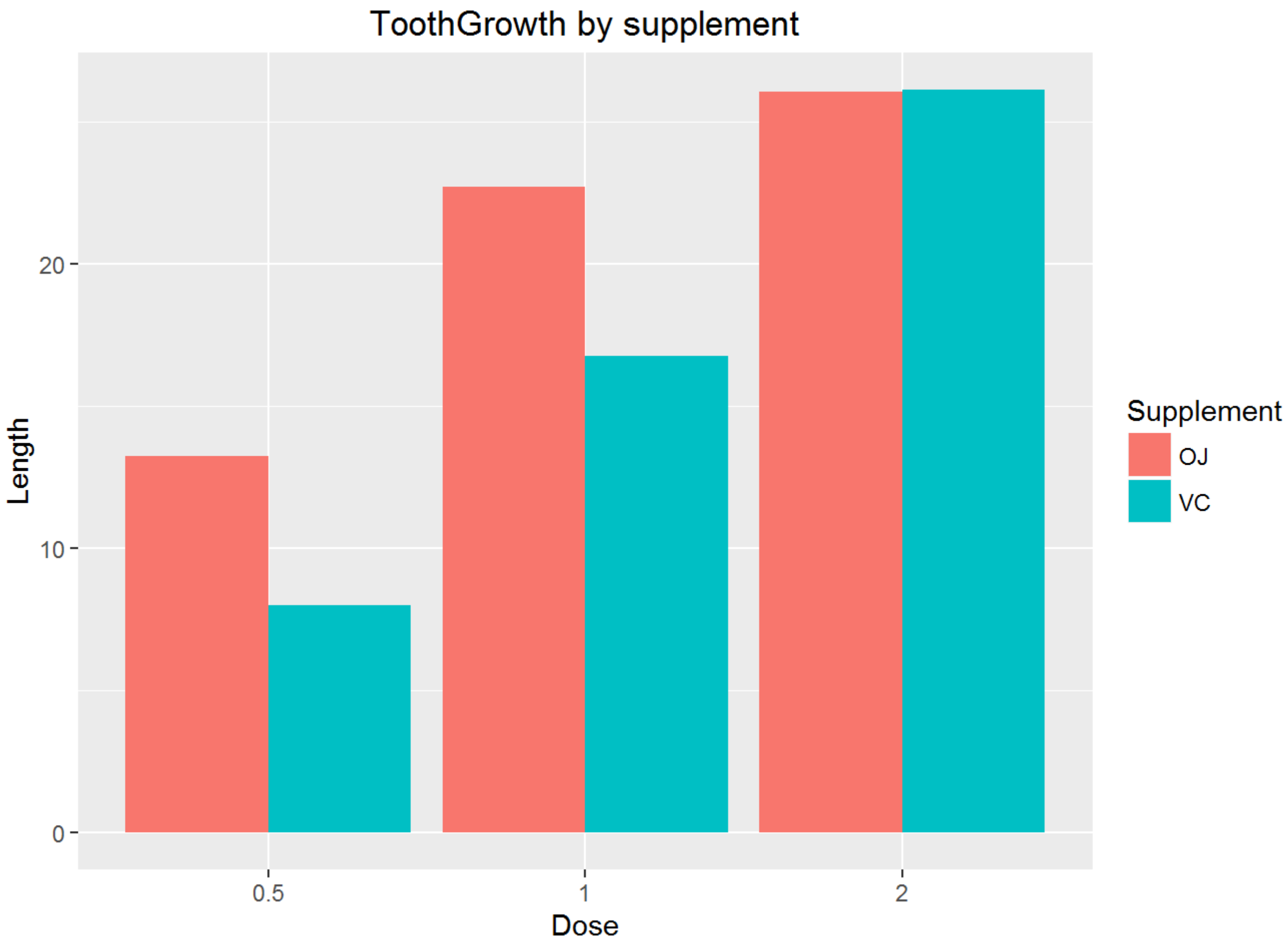
4 de outubro de 2016

For this assigment was used the dataset ToothGrowth. The CodeBook is: 1. len - numeric - Tooth length 2. supp - factor - Supplement type (VC or OJ) 3. dose - numeric - Dose in milligrams/day

1. Load the ToothGrowth data and perform some basic exploratory data analyses

A graph demonstrating the relationship of the variables:

```
data(ToothGrowth)

library(ggplot2)
library(dplyr)

filtered <- ToothGrowth %>%
    group_by(supp, dose) %>%
    summarise_each(funs(mean(len))) %>%
    as.data.frame()

filtered$dose <- as.factor(filtered$dose)

names(filtered) <- c("Supplement","Dose",'Length')

g <- ggplot(data = filtered,
    aes(x = Dose, y = Length))
g <- g + geom_bar(stat = "identity",
    aes(fill=Supplement),
    position=position_dodge())
g <- g + labs(title = "ToothGrowth by supplement")
print(g)
```



We can see that have a difference in growth of tooth relations with the supplement or dose.

2. Provide a basic summary of the data.

A little summary of the variables data:

```
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  
```

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

To calculate the confidence intervals and the p-value, we use the follow functions:

```
calculateIntervals <- function(g1, g2) {

    # Get the mean
    g1m <- mean(g1)
    g2m <- mean(g2)

    # Get the standard deviation
    g1s <- sd(g1)
    g2s <- sd(g2)

    # Get the number of each collection
    g1n <- length(g1)
    g2n <- length(g2)

    # Get the degrees of freedom
    df <- (g1s^2/g1n + g2s^2/g2n)^2 / ((g1s^2/g1n)^2/(g1n-1) + (g2s^2/g2n)^2/(g2n-1))

    # Calculate the confidence interval for 5% of both sides using T Student's
    int <- (g1m - g2m) + c(-1,1) * qt(.975, df) * sqrt(g1s^2/g1n + g2s^2/g2n)

    return(int)
}

calculatePValue <- function(g1, g2) {

    # Order to get from the smaller to greater
    if(mean(g1) > mean(g2)) {
        aux <- g1
        g1 <- g2
        g2 <- aux
    }

    # Get the mean
    g1m <- mean(g1)
    g2m <- mean(g2)

    # Get the standard deviation
    g1s <- sd(g1)
    g2s <- sd(g2)

    # Get the number of each collection
    g1n <- length(g1)
    g2n <- length(g2)

    # Get the degrees of freedom
    df <- (g1s^2/g1n + g2s^2/g2n)^2 / ((g1s^2/g1n)^2/(g1n-1) + (g2s^2/g2n)^2/(g2n-1))

    # Get T statistic
    t <- (g1m - g2m) / sqrt(g1s^2/g1n + g2s^2/g2n)

    # Return the p-value
    return(2 * pt(t, df))

}
```

Classifing by supplement

The confidence intervals for growth with supplements are:

```
g1 <- ToothGrowth$len[ToothGrowth$supp == "OJ"]
g2 <- ToothGrowth$len[ToothGrowth$supp == "VC"]
calculateIntervals(g1, g2)
```

```
## [1] -0.1710156  7.5710156
```

And the p-value is:

```
g1 <- ToothGrowth$len[ToothGrowth$supp == "OJ"]
g2 <- ToothGrowth$len[ToothGrowth$supp == "VC"]
calculatePValue(g1, g2)
```

```
## [1] 0.06063451
```

Classifing by dose of 0.5 and 1.0

The confidence intervals for dose with 0.5 and 1.0 are:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
g2 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
calculateIntervals(g1, g2)
```

```
## [1] -11.983781  -6.276219
```

And the p-value is:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
g2 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
calculatePValue(g1, g2)
```

```
## [1] 1.268301e-07
```

Classifing by dose of 0.5 and 2.0

The confidence intervals for dose with 0.5 and 2.0 are:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
g2 <- ToothGrowth$len[ToothGrowth$dose == 2.0]
calculateIntervals(g1, g2)
```

```
## [1] -18.15617 -12.83383
```

And the p-value is:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
g2 <- ToothGrowth$len[ToothGrowth$dose == 2.0]
calculatePValue(g1, g2)
```

```
## [1] 4.397525e-14
```

Classifing by dose of 1.0 and 2.0

The confidence intervals for dose with 1.0 and 2.0 are:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
g2 <- ToothGrowth$len[ToothGrowth$dose == 2.0]
calculateIntervals(g1, g2)
```

```
## [1] -8.996481 -3.733519
```

And the p-value is:

```
g1 <- ToothGrowth$len[ToothGrowth$dose == 1.0]
g2 <- ToothGrowth$len[ToothGrowth$dose == 2.0]
calculatePValue(g1, g2)
```

```
## [1] 1.90643e-05
```

Conclusions

For supplement the confidence interval of 5% the p-value was 6%, so the difference by supplement is not significant to consider one supplement instead of the other.

For dose, the p-values for dose differences are most significative and the dose should be considered. When we increase the dose, the teeth has greater growth rates.