Investigation of ToothGrowth Dataset

Course Project for **Statistical Inference** classes on **Coursera**

Igor Goltsov riversy@gmail.com

This work is my own investigation of *ToothGrowth* dataset that describes effect of Vitamin C on tooth growth in Guinea Pigs. There are two methods of Vitamin C suppliement: Orange Juice and Ascorbic Acid. And there are three doses of Vitamin C applied to pigs: 0.5, 1 and 2 miligrams.

1. Exporatory Analisys

First of all I woul like to load dataset and have a brief look onto data.

```
library(datasets)
data("ToothGrowth")
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
```

On a brief look we have 6 groups of data due to different variants of suppliement. I would like to create additional column *group* what will help me to compare these groups. I will use **dplyr** package to do that.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

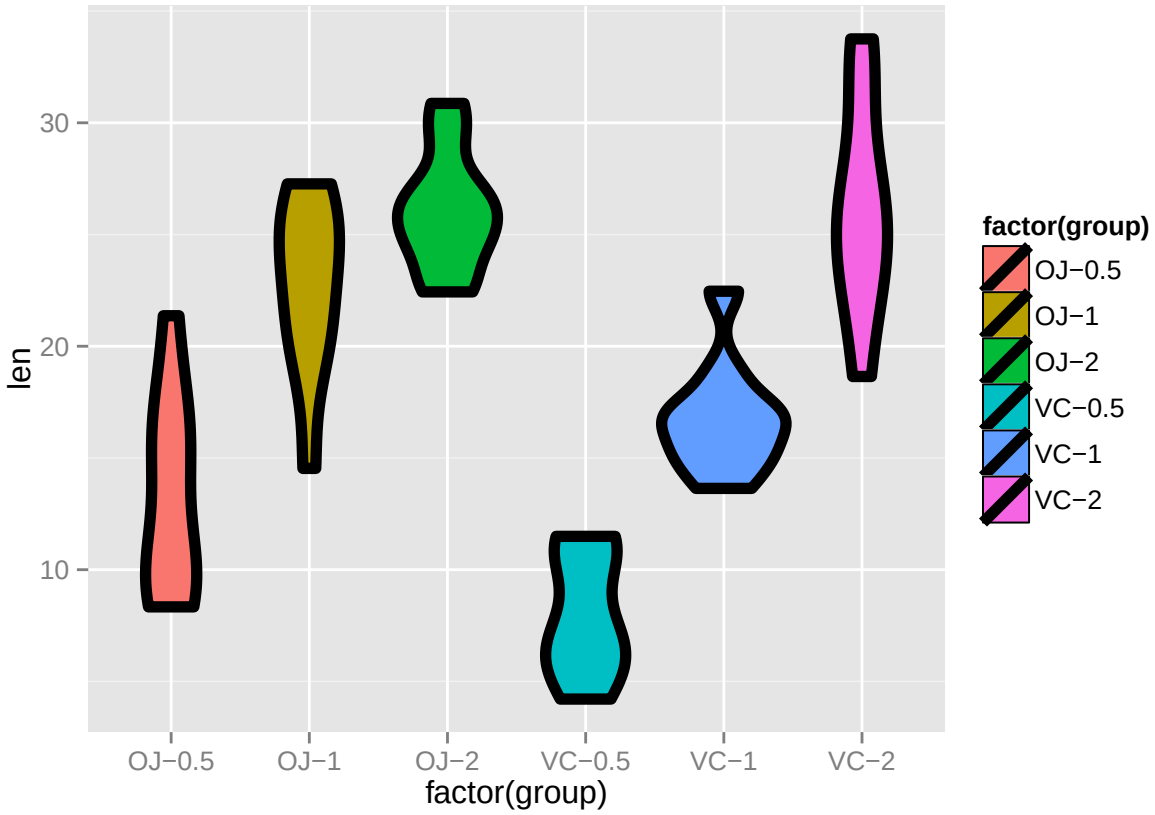
```
tooth_growth_grouped <-
    mutate(ToothGrowth, group = paste(supp, dose, sep = "-")) %>%
    select(len, group)

head(tooth_growth_grouped)
```

```
##    len  group
## 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
```

We have grouped data and I would like to visualize the density of values for each of group.

```
library(ggplot2)
ggplot(
        tooth_growth_grouped,
        aes(x = factor(group), y = len, fill = factor(group))
    ) +
    geom_violin(col = "black", size = 2)
```



Let's assume that the greater length of tooth is better for us. We may see the influence of Ascorbic Acid have a greatet effect to lenght of tooth then the use Oragnge Juice. And we also may see that the greater value of Vitamin C in Ascorbic Acid has greater influence then less value of Vitamin C.

2. Compare Ascorbic Acid quantity influence

I would like to compare two groups *VC-0.5* and *VC-2* as more outstanding groups to the minimum and to the maximum of lengths. Suppose, 2mg of Vitamin C that supplied with Ascorbic Acid has greater effect for tooth length then 0.5mg of Vitamin C that was supplied same way. I will use T-Test to check this hypothesis.

```r
vc0.5 <- tooth_growth_grouped$len[tooth_growth_grouped$group == "VC-0.5"]
vc2 <- tooth_growth_grouped$len[tooth_growth_grouped$group == "VC-2"]

t.test(vc2, vc0.5, var.equal = FALSE, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc2 and vc0.5
## t = 10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  14.41849 21.90151
## sample estimates:
## mean of x mean of y
##     26.14      7.98
```

As we may see, T-Test confidence interval is placed above zero so it's mean that our hypothesis is true. We may also may see that **mean** of *VC-2* group is much greater then **mean** of *VC-0.5* group that also confirm our suggestion.