Statistical Inference

Alexander Klonov

07 09 2020

Introduction

This task requires to do some explanatory analysis and make conclusions about data we have. The data for task is ToothGrowth

Explanotory analysis

First we need to load data:

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

Look for dimensions and collum names:

```
dim(ToothGrowth)
```

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

```
names(ToothGrowth)
```

```
## [1] "len"  "supp" "dose"
```

```
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
```

And finally let's look through summary

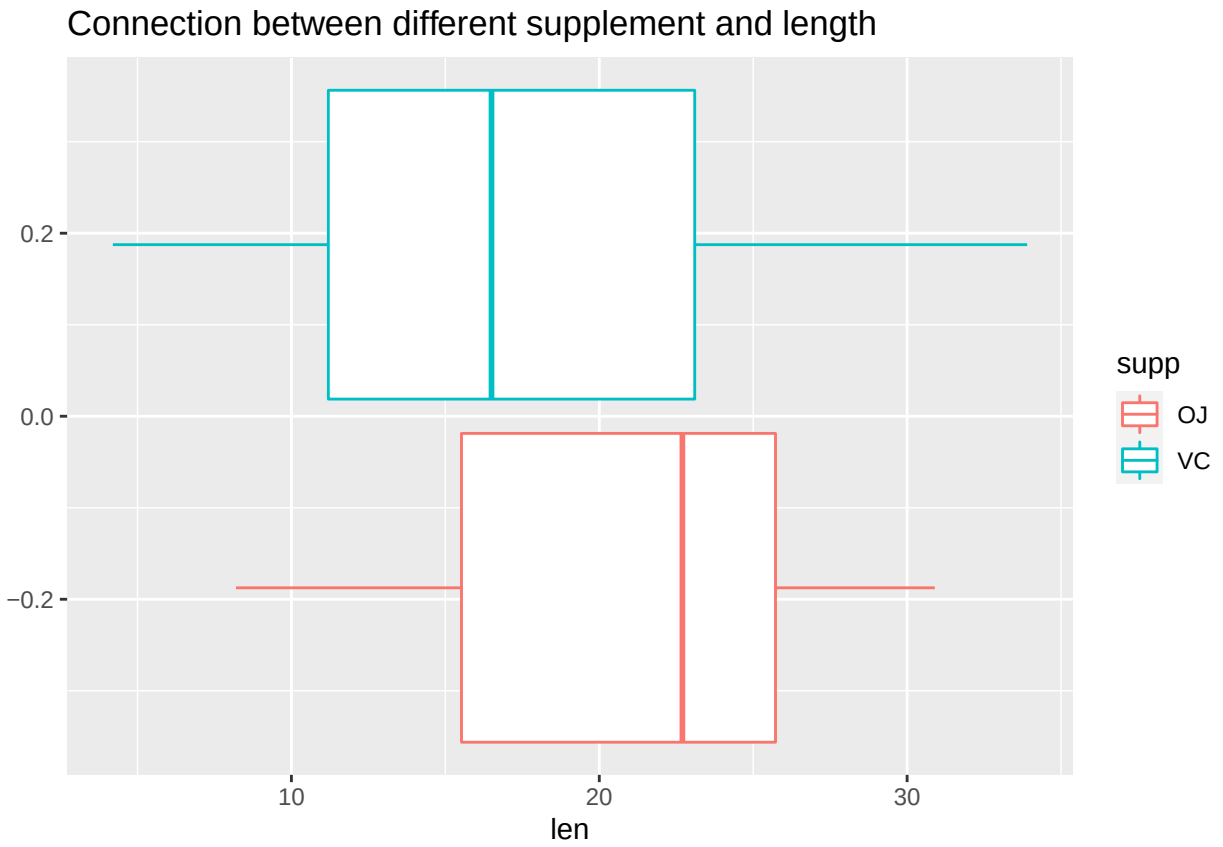
```
summary(ToothGrowth)
```

```
##       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
```

Quick graphs

To finish explanatory analyses, we need to build up a couple graphs. First will be a graph showing conection between supplement type and teeth length.

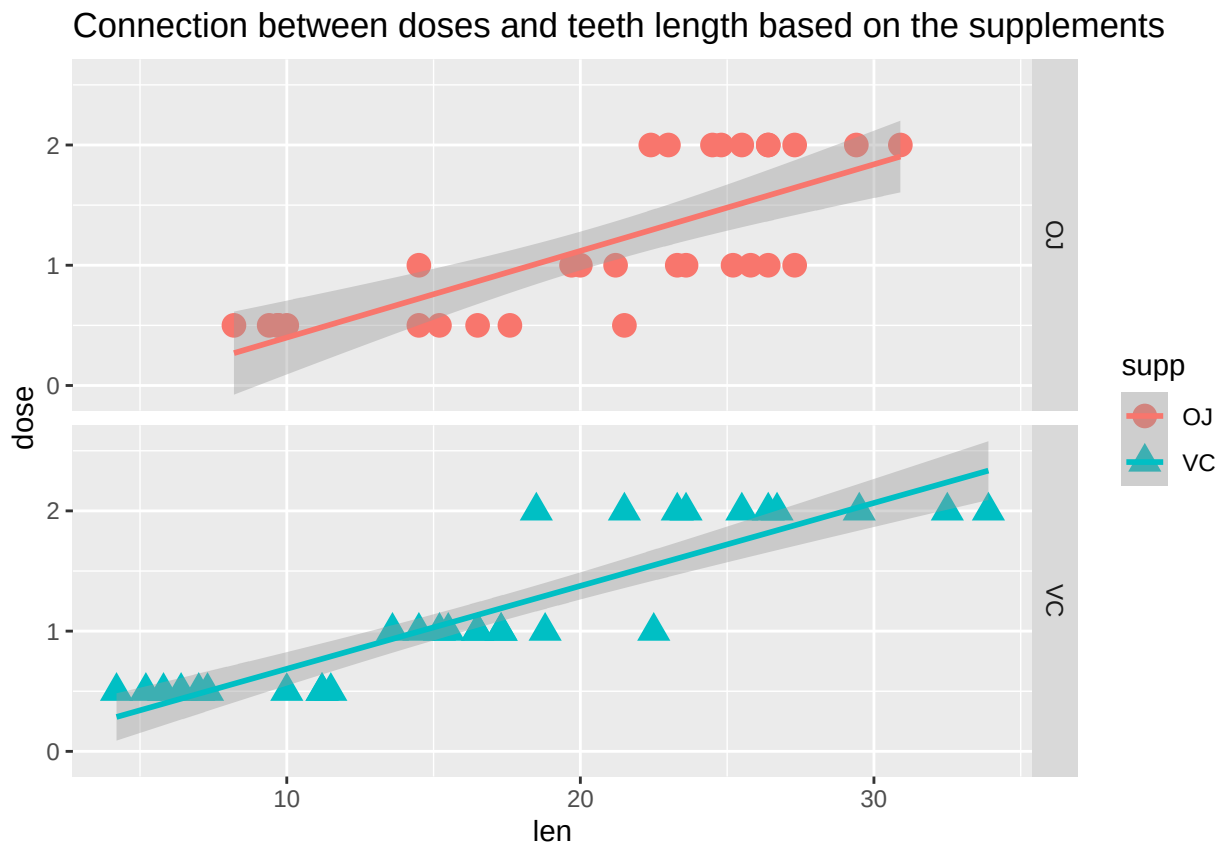
```
g<-ggplot(data = ToothGrowth, aes(len, colour = supp))+geom_boxplot()
g+ggtitle("Connection between different supplement and length")
```



And then let's add more data to this graph to see if there any connection between dose and teeth length.

```
g<-ggplot(data= ToothGrowth, aes(len, dose, shape = supp, col = supp))+
  geom_point(size = 4)+
  facet_grid(supp~.)
g+
  ggtitle("Connection between doses and teeth length based on the supplements")+
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Statistical inference

Let's assume that type of the supplement doesn't affect teeth length. To reject or fail to reject this theory let's conduct a t test.

```
Ojlen<-ToothGrowth[1:30,1] ## length by oj
Vclen<-ToothGrowth[31:60,1] ## length by Vc

t.test(Ojlen, Vclen, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  Ojlen and Vclen
## t = -3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.991341 -1.408659
## sample estimates:
## mean of the differences
##                    -3.7
```

Based on this test means in two supplements are not equal, as 95 percent confidence interval doesn't include 0 and t value is quite high.

Let's do the same, only based on suplement dose:

```
dose05<-subset(ToothGrowth, dose == 0.5)[,1]
dose1<-subset(ToothGrowth, dose == 1)[,1]
dose2<-subset(ToothGrowth, dose == 2)[,1]


t.test(dose2,dose1, paired = TRUE)$conf
```

```
## [1] 3.471814 9.258186
## attr(,"conf.level")
## [1] 0.95
```

Doubled dozes and normal appears to differ.

```
t.test(dose2,dose05,paired = TRUE)$conf
```

```
## [1] 12.6228 18.3672
## attr(,"conf.level")
## [1] 0.95
```

Difference between halfed and doubled dozes even stronger.

```
t.test(dose1,dose05,paired = TRUE)$conf
```

```
## [1]  6.387121 11.872879
## attr(,"conf.level")
## [1] 0.95
```

And normal dozes and small doesn't appears to be same.

Conslusion

We made a assumptions (null hypothesizes), that teeth length doesn't based on type of supplement and its dose. However, t test showed, that all these hypothesizes were reject. Based on this information, we can assume that teeth length does base on the supplement type and its doze.