Part 2: Basic Inferential Data Analysis Instructions

Tomislav Kralj

26/08/2022

Abstract

The Tooth Growth data is being considered. We investigate how the supplement type and dose effect the tooth growth.

Introduction and the basic exploratory data analyses

Let's load the ToothGrowth data and take a look at the dataset. We can use `help("ToothGrowth")` to see the documentation. Three columns are present:

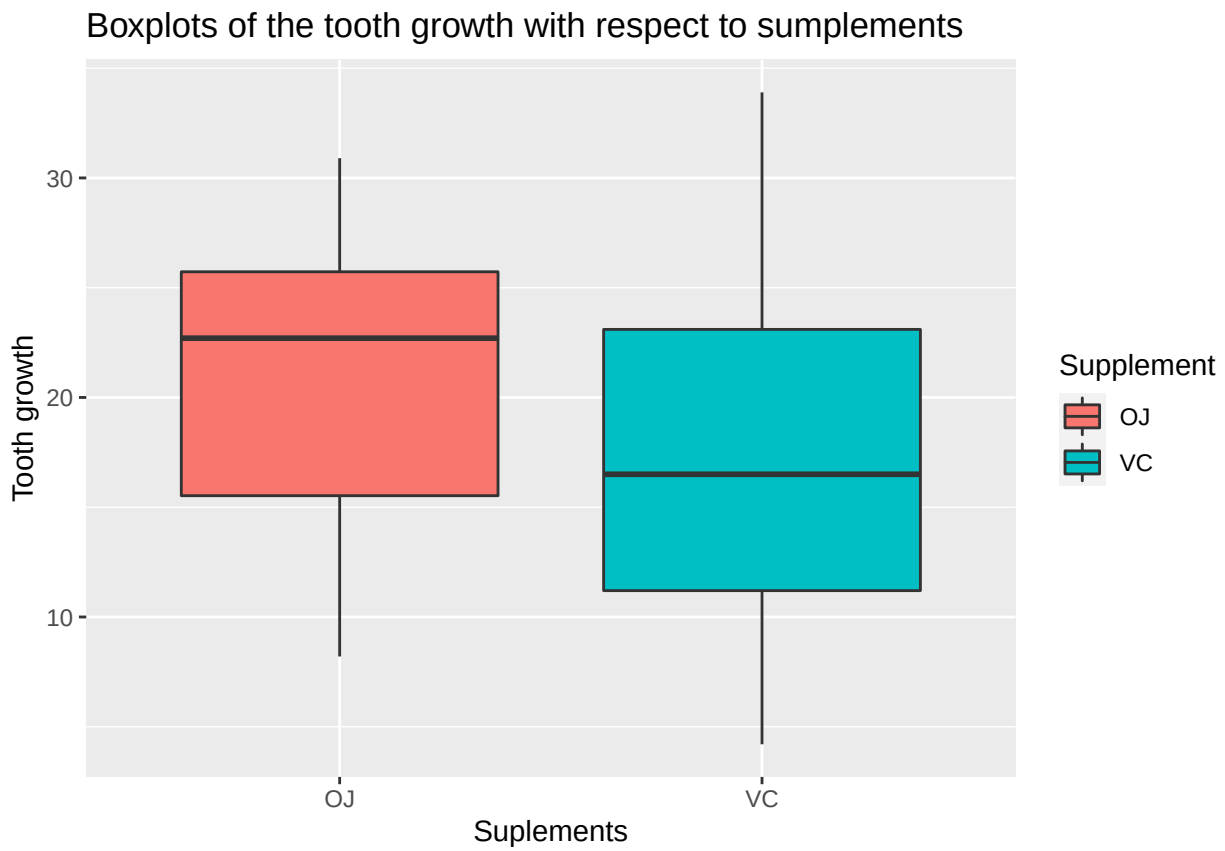
- `len` - numeric vector containing the tooth lengths,
- `supp` - factor vector (two levels) that gives the supplement type,
- `dose` - numeric vector, dose of the corresponding sumplement in milligrams per day.

```
library(datasets)
data <- datasets::ToothGrowth
str(data)
```

```
## '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 ...
```

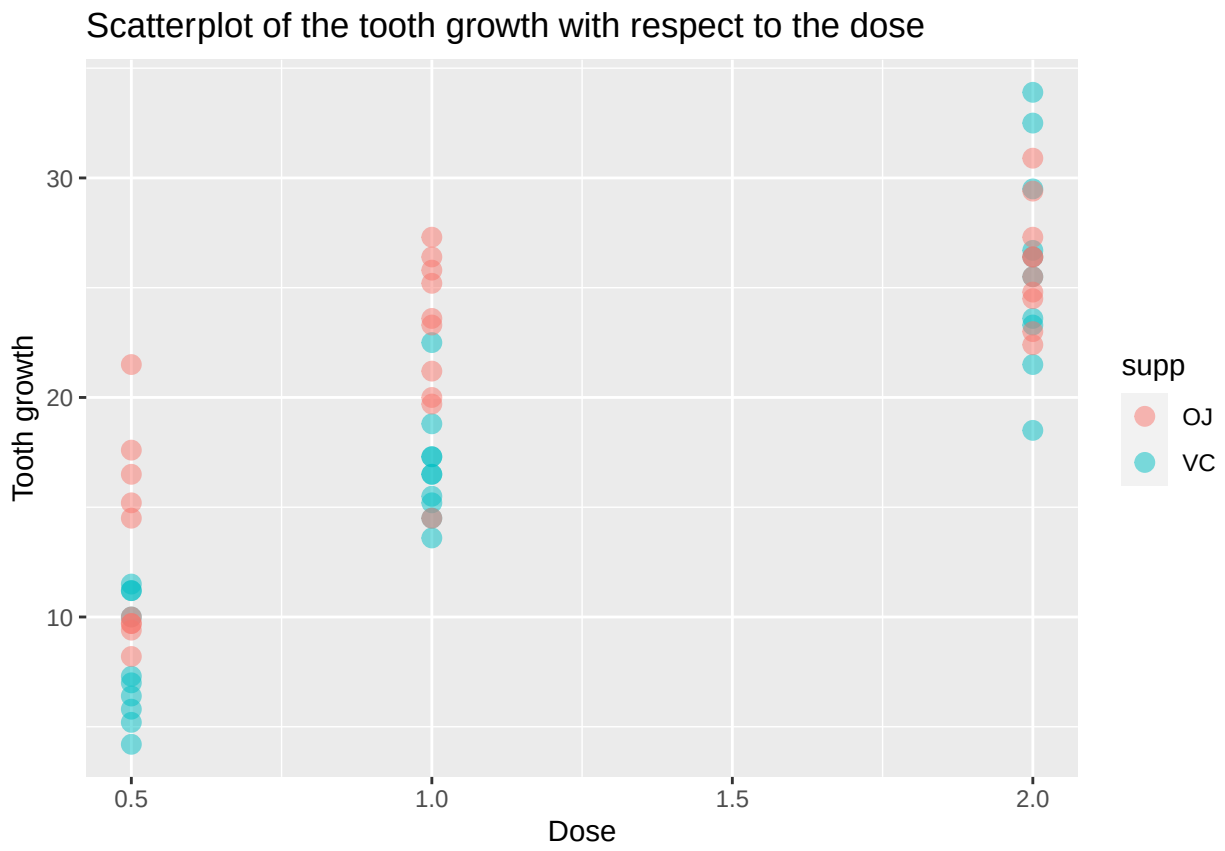
Now, let's do some plots. Firstly, let's consider the tooth growth in the relation with the type of supplement.

```
ggplot(data = data, mapping = aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp)) +
  labs(title = "Boxplots of the tooth growth with respect to sumplements") +
  xlab("Suplements") +
  ylab("Tooth growth") +
  scale_fill_discrete(name="Supplement")
```



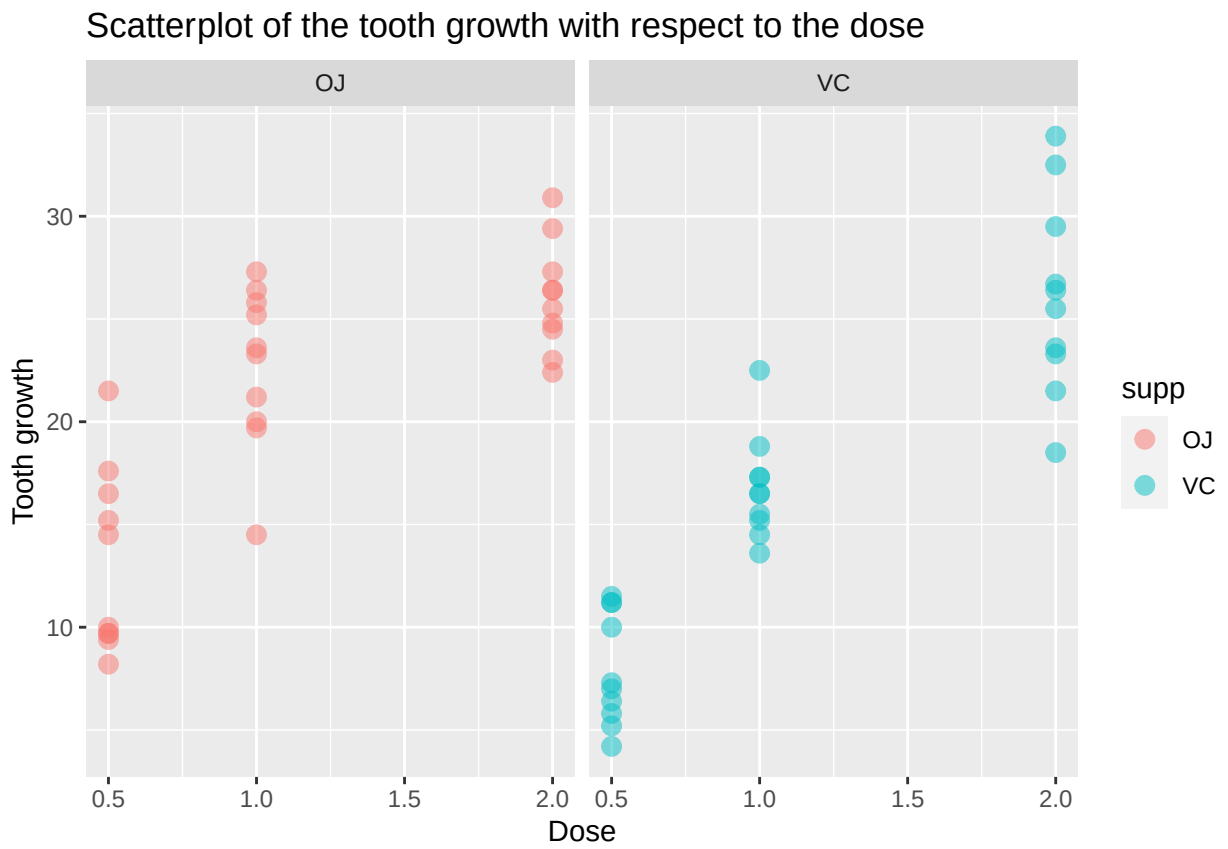
We can see that the `OJ` sumplement yields the larger tooth growth than the `VC` supplement on average, but we don't know if it's a significant difference. Similarly, we can conisder the tooth growth in the relation with dose.

```
ggplot(data = data, mapping = aes(x = dose, y = len)) +
  geom_point(aes(color = supp), size = 3, alpha = 1/2) +
  labs(title = "Scatterplot of the tooth growth with respect to the dose") +
  xlab("Dose") +
  ylab("Tooth growth") +
  scale_fill_discrete(name="Supplement")
```



One can see that with the increase of dose, the tooth growth also increases. Additionally, I've added the type of the supplement as the color to check whether the type of supplement affects the growth. Alternatively, we can make the scatterplots for each type of supplement.

```
ggplot(data = data, mapping = aes(x = dose, y = len)) +
  geom_point(aes(color = supp), size = 3, alpha = 1/2) +
  facet_grid(. ~ supp) +
  labs(title = "Scatterplot of the tooth growth with respect to the dose") +
  xlab("Dose") +
  ylab("Tooth growth") +
  scale_fill_discrete(name="Supplement")
```



It appears that the tooth growth increases evenly for both type of supplements.

Hypothesis testing

Firstly, let's check whether both types of supplements are delivering same tooth growth.

```
h1 <- t.test(len ~ supp, data = data)
h1$p.value
```

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

```
h1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The p-value is greater than the significance level, so we cannot reject the null hypothesis. Therefore, there is no sufficient evidence that the tooth growth depends on the type of supplement.

Let's see what happens if we only consider the small doses. From the graph above, it appears that `OJ` is more affective than `VC` when conisdering only the dose of 0.5.

```r
h2 <- t.test(len ~ supp, data = data %>% filter(dose == 0.5), alternative = "greater")
h2$p.value
```

```
## [1] 0.003179303
```

```r
h2$conf.int
```

```
## [1] 2.34604     Inf
## attr(,"conf.level")
## [1] 0.95
```

The p-value is less that the significance level, which gives us that we can reject the null hypothesis in favor of the alternative which states that `OJ` has greater population mean.

Let's see what happens in other two levels of doses. Namely, we consider the growth when `dose == 1` and `dose == 2`, separately.

```r
t.test(len ~ supp, data = data %>% filter(dose == 1))$p.value
```

```
## [1] 0.001038376
```

```r
t.test(len ~ supp, data = data %>% filter(dose == 2))$p.value
```

```
## [1] 0.9638516
```

We can see that in the case `dose == 1`, these two supplements yield statistically different average tooth gowth, while in the case `dose == 2`, there is no statistical difference between these two supplements. Additionaly, we check that `OJ` is indeed more affective in the case `dose == 1`.

```r
t.test(len ~ supp, data = data %>% filter(dose == 1), alternative = "greater")$p.value
```

```
## [1] 0.0005191879
```

Conclusion

In the prior testing, we did not observe significant difference between types of supplements. But, when we divided the growth depending on the doses, we see that in the case of `dose == 0.5` and `dose ==1` `OJ` is more effective than `VC`, but there is no significant difference in the case of `dose == 2`.