

Course_Project_IS_part2

Dataset: The effect of Vitamin C on Tooth Growth in Guinea Pigs. The variables of the data set are len = tooth length, supp = supplement type (VC or OJ) and dose = does in milligrams per day.

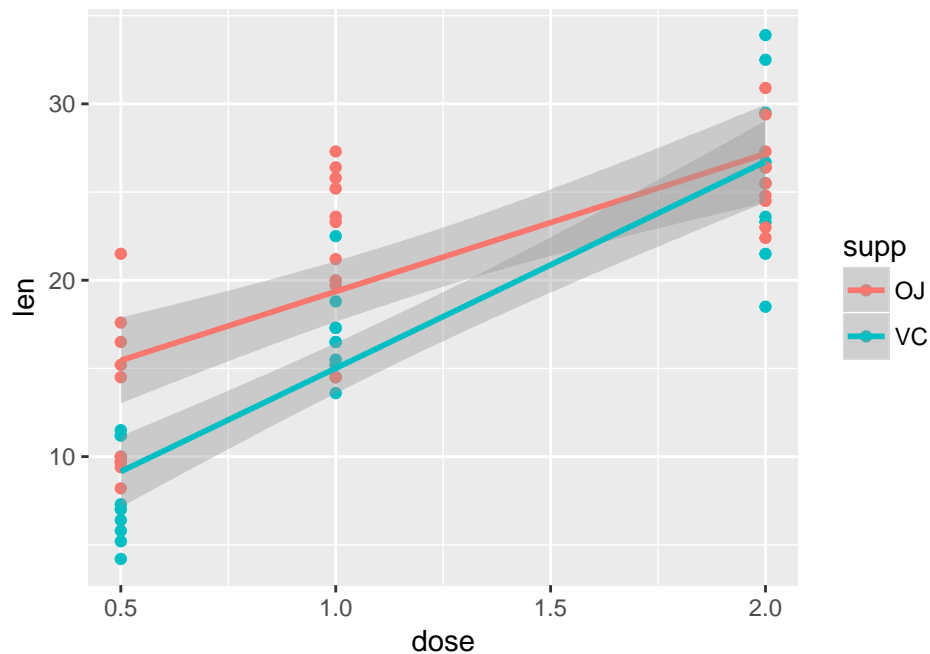
```
data("ToothGrowth")
data = ToothGrowth
```

```
summary(data)
```

```
##      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
```

Plot length against dose for every supplement.

```
g <- ggplot(data, aes(x = dose, y = len, color = supp)) +
  geom_point() + geom_smooth(method = "lm")
g
```



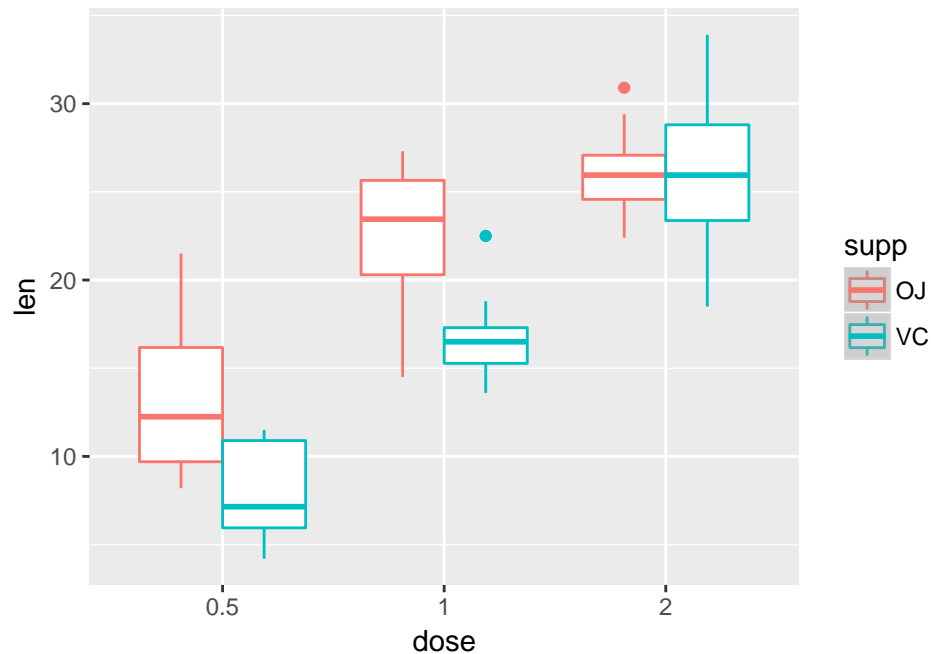
Based on this plot there are different aspects that can be investigated such as: Does the overall tooth length depend on the supplement or does the tooth length depend on the dose or does the tooth length depend on the supplement for a given dose etc. . . In order to investigate these aspects one could perform hypothesis testing.

Let us look at the last aspect: Does the tooth length depend on the supplement for a given dose.

```
data$dose = as.factor(data$dose)
```

```
g <- ggplot(data, aes(x = dose, y = len, color = supp)) +
```

```
geom_boxplot() + geom_smooth(method = "lm")
g
```



For the given problem statement the answer is quite clear for both dose = 1 and dose = 2. So let's look at the problem for dose = 0.5. Let us state the null hypothesis: The means of the tooth length are independent of supplement. The alternative hypothesis is: The means of the tooth length differ for different supplement.

```
len_p5_VC = data[data$dose == "0.5" & data$supp == "VC",]$len
len_p5_OJ = data[data$dose == "0.5" & data$supp == "OJ",]$len

t.test(len_p5_VC, len_p5_OJ, paired = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: len_p5_VC and len_p5_OJ
## 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:
## -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98      13.23
```

The result of the test lets me conclude (the small p-value < 0.01), that the tooth length is indeed different for each of the two given supplement at dose = 0.5.