Statistical Inference Course Project - Part 2

Artem Braun

Statistical Inference Course Project - Part 2

We are going to analyze the ToothGrowth data in the R datasets package and provide some conclusions in order to demonstrate inferential analysis skills. Since it is stated in the task that the code should be listed as supporting appendix, I provide report first and the code afterwards.

1. Basic exploratory data analysis

From R documentation we can get the following information:

This dataset measures the influence of Vitamin C on tooth growth in Guinea Pigs. Dataset includes observations for different combinations of 2 supplements (OJ - orange juice and VC - Vitamin C) and 3 dosages (0.5mL, 1mL and 2mL). Using basic 'summary' function we can derive that there are 10 observations in dataset for each combination of dose and supplement, totalling 60 observations.

One of the best ways for the first look at the data is box plot. Plot 1 in appendix shows the range of length for each combination of dose and supplement. Looking at the plot, we can infer that the higher the supplement dosage - the higher the tooth length. Additionally, it looks like the mean for supplement VC is less than for OJ for doses 0.5 and 1, but for the dose 2 the difference is not so sound. To reaffirm our inference we can look at the aggregated means in Table 1. We see the same trend.

2. Hypothesis test to compare tooth growth by supplement and dose

Based on our initial inferential analysis we can define the following hypothesis: tooth lenths do not depend on different supplements. In other words: does the tooth lenth statistically different between the VS and OJ supplements for each dosage.

We will use t-test to test this hypothesis. We should do this test for each of dosages: 0.5, 1 and 2. If the p-values are less than 0.05, then we can determine our inferences with 95% confidence.

The p-values for dosages 0.5 and 1 were quite small, whereas for dosage 2 p-value was significantly larger than 0.05, since the mean values are close together.

3. Assumptions and coclusions

Based on provided t-tests we can conclude that for dosages up to 1mL orange juice affect teeth growth considerably more than Vitamin C. For dosages more that 1mL supplements orange juice and Vitamin C affect the teeth growth quite similar.

Supporting appendix material (code and figures)

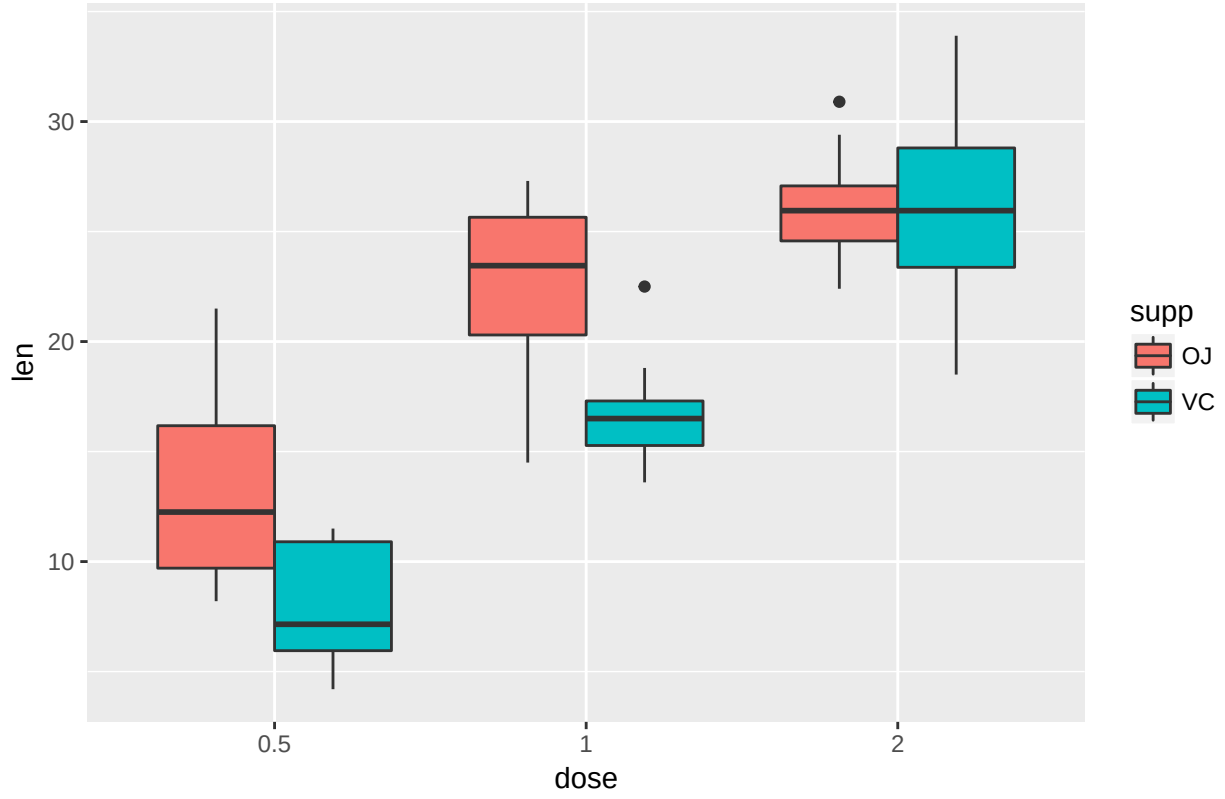
```
library(dplyr)
library(ggplot2)

Raw_data <- ToothGrowth
summary(Raw_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
```

```
Raw_data$dose <- factor(Raw_data$dose)
ggplot(Raw_data) +
        geom_boxplot() +
        aes(x=dose,y=len, fill= supp) +
        labs(title= "Plot 1. Tooth growth split by supplement and dose")
```

Plot 1. Tooth growth split by supplement and dose




```
Lenth_means <- aggregate(len ~ ., data = Raw_data, mean)
Lenth_means <- split(Lenth_means, Lenth_means$dose)
print("Table 1. Means of tooth lengths")
```

```
## [1] "Table 1. Means of tooth lengths"
```

```
print(Lenth_means)
```

```
## $`0.5`
##   supp dose   len
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
##
## $`1`
##   supp dose   len
## 3   OJ    1 22.70
## 4   VC    1 16.77
##
## $`2`
##   supp dose   len
## 5   OJ    2 26.06
## 6   VC    2 26.14
```

```
t.test(len~supp,Raw_data[Raw_data$dose=="0.5",])
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by supp
## 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:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp,Raw_data[Raw_data$dose=="1",])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70             16.77
```

```
t.test(len~supp,Raw_data[Raw_data$dose=="2",])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06             26.14
```