Statistical Inference Project Part 2

Moisés R. Santos

Basic inferential data analysis

Purpose

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Load library

```
library("dplyr")
library("ggplot2")
library("patchwork")
```

Load the ToothGrowth data and perform some basic exploratory data analyses

```
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
```

```
str(ToothGrowth)
```

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

```
summary(ToothGrowth)
```

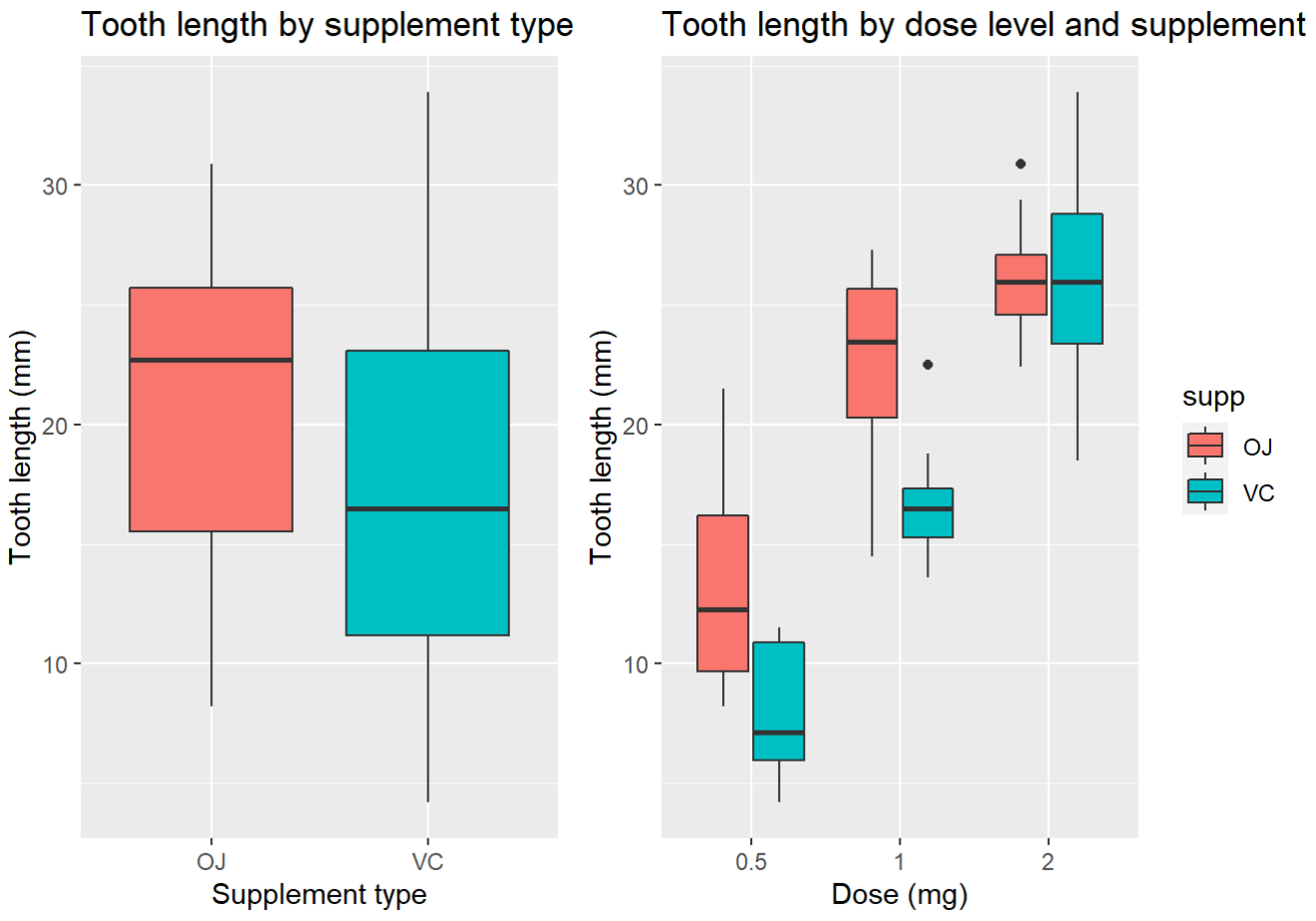
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Provide a basic summary of the data.

```
g_len_by_supp <- ggplot(ToothGrowth, aes(supp, len)) +
  geom_boxplot(aes(fill = supp), show.legend = F) +
  xlab('Supplement type') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by supplement type')

g_len_by_dose <- ggplot(ToothGrowth, aes(factor(dose), len)) +
  geom_boxplot(aes(fill = supp)) +
  xlab('Dose (mg)') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by dose level and supplement type')

(g_len_by_supp | g_len_by_dose)
```



Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

First, let's check if there is a significant difference between the 2 types of supplements: orange juice and vitamin C. In this t test, it was considered that the variances are different. 95% confidence was considered.

```
t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE)
t_diff_supp
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

As can be seen, considering 95% confidence the null hypothesis cannot be rejected, since the p-value is greater than 0.05. So we have no evidence that there is a significant difference between the two supplements.

Now, let's see if there is a significant difference between supplements at different dosages.

dosage = 0.5

```
t_diff_supp_0.5 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 0.5))
t_diff_supp_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
```

dosage = 1

```
t_diff_supp_1 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 1))
t_diff_supp_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
```

dosage = 2

```
t_diff_supp_2 <- t.test(len ~ supp, data = subset(ToothGrowth, dose == 2))
t_diff_supp_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
```

Conclusions

Considering 95% confidence, from the t-tests it was possible to conclude that in general there is no significant difference between the different supplements. However, when analyzing the dosage level, there is a significant difference for 0.5 and 1 mg. It is not possible to say that for 2 mg the supplements have different growth. p-value <0.05 was considered relevant.