Hypothesis Testing Using the T Test

Andres Camilo Zuñiga Gonzalez

23/5/2020

As a first step, it is necessary to load the necessary packages and set the working directory.

```r
setwd('./Statistical_Inference')

library(ggplot2) #Plotting system
library(cowplot) #Panel for ggplot2
```

Exploratory Data Analysis

First we load the data and show the structure and summary of the variables.

```r
data("ToothGrowth")
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 ...
```

```r
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
```

Then, we build some visualizations to see the distribution of the variables and the relationship between them.

```r
main <-ggplot(ToothGrowth, aes(x = factor(dose), y = len)) +
    geom_boxplot(aes(fill = supp)) +
    labs(x = 'Dose', y = 'Length', fill = 'Supplement') +
    theme_light() + theme(legend.position = 'none')

dist <- ggplot(ToothGrowth, aes(x = len, fill = supp)) +
    geom_histogram(binwidth = 1) + geom_density(aes(y = ..count.., fill = NULL)) +
    facet_grid(vars(dose), vars(supp)) +
    labs(x = 'Length', y = 'Frequency', fill = 'Supplement') +
    theme_light()

plot_grid(main, dist)
```
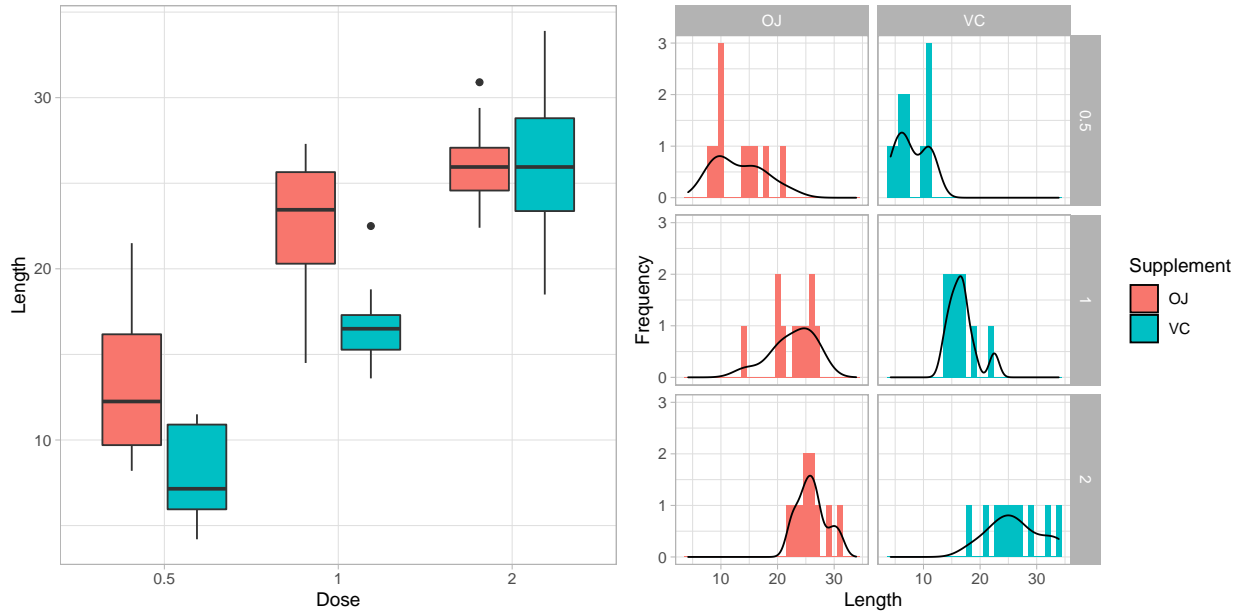
Supplement Effect on Tooth Length Hypothesis Testing

In the first analysis I will compare the mean of the tooth length according to the supplement each study subject took.

$$H_0 : \mu_{OJ} = \mu_{VC}$$

$$H_a : \mu_{OJ} \neq \mu_{VC}$$

I want to check if there is a significant difference between observation with both types of supplements. For this reason, I need a t test for independant samples.

```
t_test <- t.test(len ~ supp, data = ToothGrowth)
```

According to this test, with a p-value of **0.0606345** an a confidence interval of [**-0.1710156, 7.5710156**]. According to these results, it is not possible to reject the null hypothesis, therefore, the supplement does not have an effect on tooth growth. In addition, the confidenci interval contains 0, which adds to the conclusion of not rejecting the null hypothesis.

Dose Effect on Tooth Length Hypothesis Testing

For this test, it is necessary to make subsets for each dose. For any case these would be the hypothesis:

$$H_0 : \mu_a = \mu_b$$

$$H_a : \mu_a \neq \mu_b,$$

for a and b being two different types of doses in $\{0.5, 1, 2\}$.

For this analysis, I will create a function to subset the dataset, removing one of the doses and comparing the mean of the other two in a t test.

```
t_test_dose <- function(dose1, dose2) {
    df <- subset(ToothGrowth, dose %in% c(dose1, dose2))
    return(t.test(len ~ dose, data = df))
}
```

```
dose0.5_vs_1 <- t_test_dose(0.5, 1)
dose1_vs_2 <- t_test_dose(1, 2)
dose0.5_vs_2 <- t_test_dose(0.5, 2)
```

The following table show the p-value and the confidence intervals for each of the tests:

Comparison	p-value	Confidence Interval
Dose 0.5 vs Dose 1	0.0000001	[-11.9837813, -6.2762187]
Dose 1 vs Dose 2	0.0000191	[-8.9964805, -3.7335195]
Dose 0.5 vs Dose 2	0	[-18.1561665, -12.8338335]

According to these results, for every type of doses, it is possible to reject the null hypothesis. In all three cases, the confidence interval do not contain zero, adding evidence to the statistical conclusion. Given these results, the dose does have an effect on tooth length.

It is worth noting that in order to do a t test, samples must be random, and must be representative of the population, and the variances of the two groups to be compared must be equal.