Statistical inference - Course Project: Part 2

- Author: Anderson Hitoshi Uyekita
- Date: Monday, 04 July 2022

Synopsis

This project aims to analyze the Tooth Growth database using confidence intervals and/or tests. This dataset has 60 observations and 3 variables, and a summary was provided with a brief of exploratory analysis. As a results of this project, supplement type has no effect on tooth growth and increasing the dose level leads to increased tooth growth.

Requeriments, Settings, and Load Data

- Requirements to reproduce this exercise: `ggplot2`, `dplyr`, and `datasets`.
- Make a copy of the original dataset and converting into a `dplyr` table.

```
dataset_tg <- ToothGrowth
dataset_tg <- tbl_df(dataset_tg)
```

```
## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

A brief exploratory analysis of the dataset

A data frame with 60 observations on 3 variables.

How is the data structure:

```
## tibble [60 x 3] (S3: tbl_df/tbl/data.frame)
##  $ len : num [1:60] 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 [1:60] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Note: The experiment counts with two kinds of supplements (VC or JC) and there are 3 dose variation (0.5, 1.0, and 2.0 mg/day).

To synthesize all information we will use one graph.

```
ggplot(aes(x=supp, y=len), data=dataset_tg) + geom_boxplot(aes(fill=supp)) + xlab("Supplement type") +yl
```
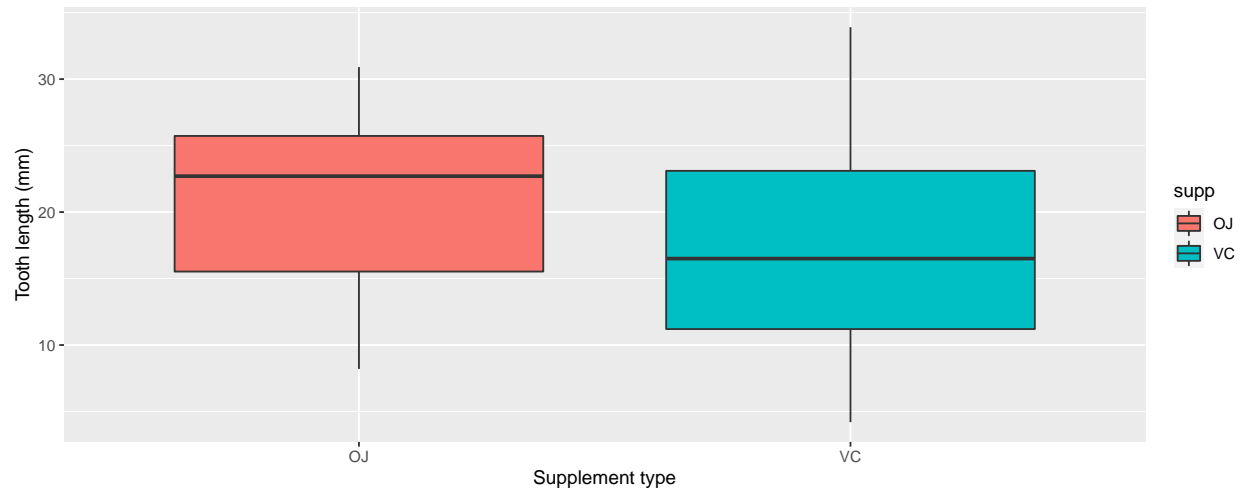
Provide a basic summary of the data

Summary of this dataset:

```
##       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
```

For futher more information about each variable, please read this document.

Compare tooth growth by supplement and dose

There are more than one comparison of tooth growth by supplement (`OJ` and `VC`) and dose. Thus, to turn this study much clearly we divided this section into 3 parts: Comparison between supplements, 1mg and 0.5 mg dose, and 2 and 1 mg dose.

Test 1: Growth Tooth Differences between supplements OJ and VC   We are testing if exist some differences between those supplements. It means, we are looking for a p value greater then 0.05. Thus, we need to assume two hypotesis: `Ho` equals means and `H1` means are differents.

First of all, we need to check the `len` variance between `OJ` and `VC` supplement.

Those variance are far different so the `var.equal` should be set to `FALSE` (`Len` Variance using OJ =43.6334368 and `Len` Variance using VC = 68.3272299). Now, we can use de `t.test` to compare the supplements performances are the same.

```
t.test(dataset_tg_OJ$len, dataset_tg_VC$len, paired = FALSE, var.equal = FALSE)
```

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

```
## data:  dataset_tg_OJ$len and dataset_tg_VC$len
## 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 of x mean of y
##  20.66333  16.96333
```

The p-value of the test is 0.06. It means we do not have evidence to reject the null hypothesis. Supplement types seems to have no impact on Tooth growth.

Test 2: Growth Tooth Differences by dosages 2 and 1 mg/day   For this test we define the `Ho` as the null hypotheses of equal means between the two groups, versus the alternative hypothesis (`H1`) that the two means are different.

```
t.test(filter(dataset_tg,dose==2)$len, filter(dataset_tg,dose==1)$len, paired = FALSE, var.equal = TRUE]
```

```
##
##   Two Sample t-test
##
## data:  filter(dataset_tg, dose == 2)$len and filter(dataset_tg, dose == 1)$len
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

As results of the t.test we have enough evidence to reject the null hypothesis (Ho). It means if I increase the dosage from 1mg to 2mg creates an positive effect on tooth length.

Test 3: Growth Tooth Differences between supplements OJ and VC   For this test we define the `Ho` as the null hypotheses of equal means between the two groups, versus the alternative hypothesis (`H1`) that the two means are different.

```
t.test(filter(dataset_tg,dose==1)$len, filter(dataset_tg,dose==0.5)$len, paired = FALSE, var.equal = TRU
```

```
##
##   Two Sample t-test
##
## data:  filter(dataset_tg, dose == 1)$len and filter(dataset_tg, dose == 0.5)$len
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

As results of the t.test we have enough evidence to reject the null hypothesis (Ho). It means if I increase the dosage from 0.5mg to 1mg creates an positive effect on tooth length.

Conclusions

- By the Test 1 we can conclude that supplement has no effect on tooth growth.

- By the Test 2 and Test 3 we can conclude that increasing the dose level leads to increased tooth growth.

Assumptions

- For t-tests regarding tooth length per dosage level, the variances are assumed to be equal for the three combinations of the two groups being compared.