# Analysis of the 'ToothGrowth' dataset

Feroz Mohamed Hatha

22/09/2020

```
knitr::opts_chunk$set(echo = TRUE)
```

## Overview

The `ToothGrowth` dataset consists of the length of the teeth recorded for 10 Guinea pigs each of whom take 3 doses (0.5 mg, 1 mg, and 2 mg) of 2 Vitamin C supplements (Orange juice and Ascorbic acid). Thus, there are a total of 60 observations (rows) and 3 variables (columns). The 3 variables are:

- len: Tooth length (numerical)
- supp: Vitamin C supplement ("OJ" or "VC") (factor)
- dose: Dosage in mg (0.5, 1, or 2) (numerical)

```
# Load the required packages
library(datasets)
library(ggplot2)
library(dplyr)
```

### 1. Loading the data + basic exploratory data analyses

```
# Load the ToothGrowth dataset
data("ToothGrowth")

# Look at the structure of the data
str(ToothGrowth)
```

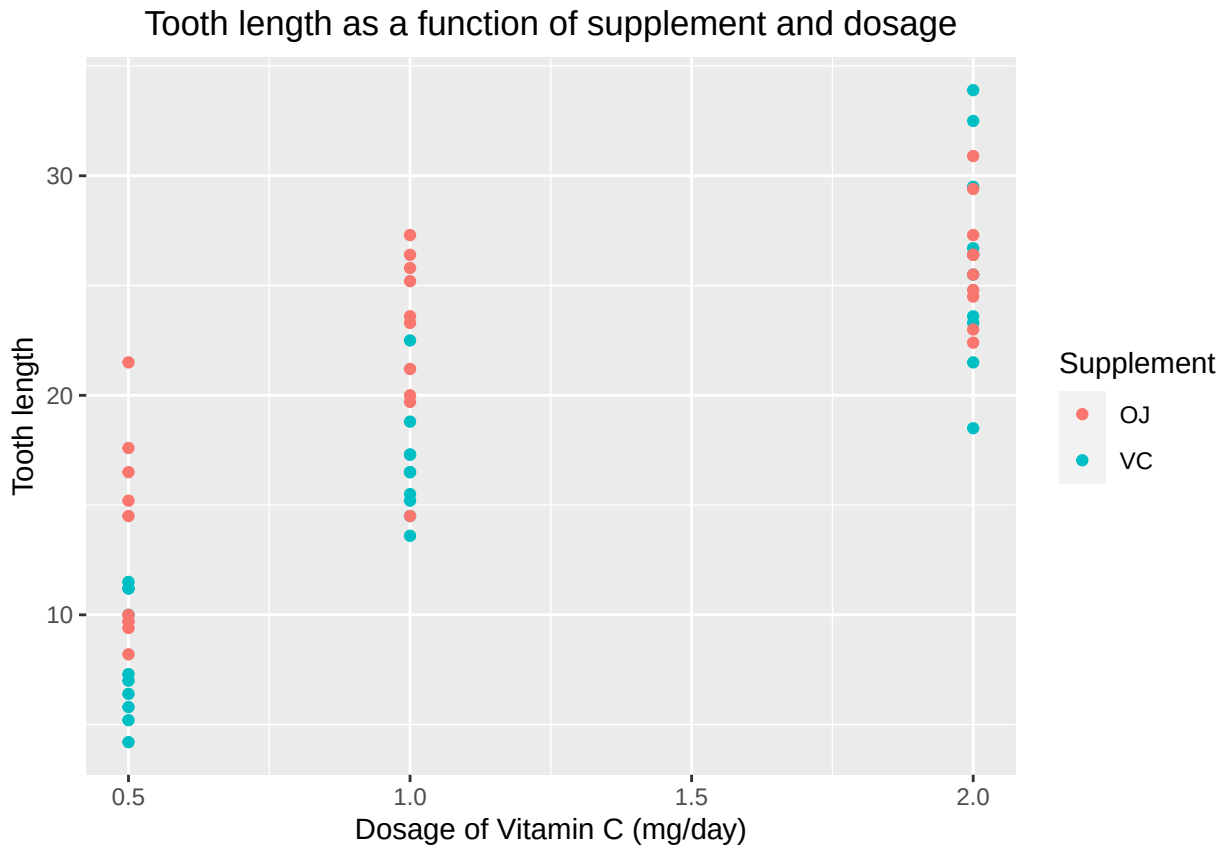
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# Look at the first few rows of the data
head(ToothGrowth)
```

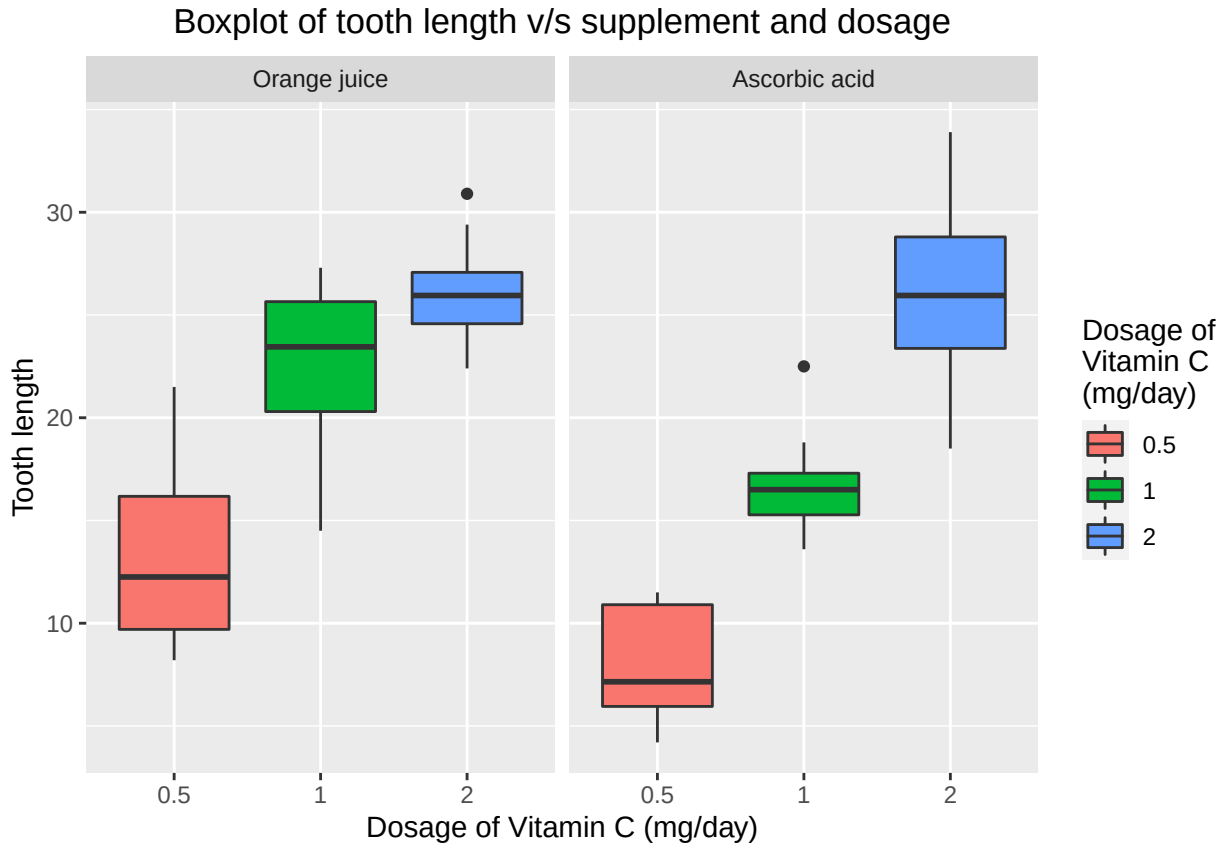
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
```

```
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
# Plot tooth length as a function of dosage and supplement
ggplot(ToothGrowth, aes(x = dose, y = len)) + geom_point(aes(color = supp)) +
  labs(title = "Tooth length as a function of supplement and dosage",
       x = "Dosage of Vitamin C (mg/day)", y = "Tooth length") +
  theme(plot.title = element_text(hjust = 0.5)) + labs(color = "Supplement")
```




```
# Make a boxplot of tooth length as a function of dosage for each supplement
ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose))) +
      geom_boxplot() +
      facet_grid(.~supp, labeller = as_labeller(
            c("OJ" = "Orange juice",
              "VC" = "Ascorbic acid"))) +
      labs(title = "Boxplot of tooth length v/s supplement and dosage",
           x = "Dosage of Vitamin C (mg/day)",
           y = "Tooth length") +
      scale_fill_discrete(name = "Dosage of\nVitamin C\n(mg/day)") +
  theme(plot.title = element_text(hjust = 0.5))
```




## 2. Provide a basic summary of the data

```
# Look at the summary for each of the three variables
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
```

```
# Get the summary for each Vitamin C supplement
ToothGrowth %>%
  group_by(supp) %>%
  summarize(mean_len = mean(len), sd_len = sd(len), count = n()) %>%
  print()
```

```
## # A tibble: 2 x 4
##   supp  mean_len sd_len count
##   <fct>    <dbl>  <dbl> <int>
## 1 OJ        20.7   6.61    30
## 2 VC        17.0   8.27    30
```

```r
# Get the summary for each dosage
ToothGrowth %>%
  group_by(dose) %>%
  summarize(mean_len = mean(len), sd_len = sd(len), count = n()) %>%
  print()
```

```
## # A tibble: 3 x 4
##    dose mean_len sd_len count
##   <dbl>    <dbl>  <dbl> <int>
## 1   0.5     10.6   4.50    20
## 2   1       19.7   4.42    20
## 3   2       26.1   3.77    20
```

```r
# For each value pair of supplement and dosage, get the summary
ToothGrowth %>%
  group_by(supp, dose) %>%
  summarize(mean_len = mean(len), sd_len = sd(len), count = n()) %>%
  print()
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp   dose mean_len sd_len count
##   <fct> <dbl>    <dbl>  <dbl> <int>
## 1 OJ      0.5    13.2    4.46    10
## 2 OJ      1      22.7    3.91    10
## 3 OJ      2      26.1    2.66    10
## 4 VC      0.5     7.98   2.75    10
## 5 VC      1      16.8    2.52    10
## 6 VC      2      26.1    4.80    10
```

From the above summaries, it seems like tooth length is directly proportional to the dosage of the Vitamin C supplement. It also looks like Orange juice is the better supplement of the two as the mean tooth length is higher in the case of Orange juice.

## 3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supplement and dosage

```r
# Compare the two supplements by using a paired t-test
supp1 <- subset(ToothGrowth, ToothGrowth$supp == "VC")
supp2 <- subset(ToothGrowth, ToothGrowth$supp == "OJ")
t.test(supp1$len, supp2$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp1$len and supp2$len
## t = -3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -5.991341 -1.408659
## sample estimates:
## mean of the differences
##                    -3.7
```

We reject the null hypothesis as the 95% confidence interval does not contain 0. Therefore, we conclude that the two supplements differ in terms of how they affect tooth growth.

```r
# Compare the two dosages 0.5 mg and 1 mg for Ascorbic acid using a paired t-test
supp_1_low <- subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5)
supp_1_mid <- subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 1)
t.test(supp_1_low$len, supp_1_mid$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp_1_low$len and supp_1_mid$len
## t = -6.1364, df = 9, p-value = 0.0001715
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.030399  -5.549601
## sample estimates:
## mean of the differences
##                   -8.79
```

Once again, we reject the null hypothesis as the 95% confidence interval does not contain 0. We conclude that the two dosages of Ascorbic acid (0.5 mg and 1 mg) differ in terms of how they affect tooth growth.

```r
# Compare the two dosages 1 mg and 2 mg for Ascorbic acid using a paired t-test
supp_1_high <- subset(ToothGrowth, ToothGrowth$supp == "VC" & ToothGrowth$dose == 2)
t.test(supp_1_mid$len, supp_1_high$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp_1_mid$len and supp_1_high$len
## t = -5.346, df = 9, p-value = 0.0004648
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.334918  -5.405082
## sample estimates:
## mean of the differences
##                   -9.37
```

We reject the null hypothesis yet again as the 95% confidence interval does not contain 0. We conclude that the two dosages of Ascorbic acid (1 mg and 2 mg) differ in terms of how they affect tooth growth.

```r
# Compare the two dosages 0.5 mg and 2 mg for Ascorbic acid using a paired t-test
t.test(supp_1_low$len, supp_1_high$len, paired = TRUE)
```

```
##
##  Paired t-test
```

```
##
## data:  supp_1_low$len and supp_1_high$len
## t = -9.7912, df = 9, p-value = 4.264e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22.3557 -13.9643
## sample estimates:
## mean of the differences
##                  -18.16
```

As expected, we get a very low p-value (and a 95% confidence interval that does not contain 0) and therefore we conclude that the two dosages of Ascorbic acid (0.5 mg and 2 mg) very clearly differ in terms of how they affect tooth growth.

```
# Compare the two dosages 0.5 mg and 1 mg for Orange juice using a paired t-test
supp_2_low <- subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5)
supp_2_mid <- subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1)
t.test(supp_2_low$len, supp_2_mid$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp_2_low$len and supp_2_mid$len
## t = -4.1635, df = 9, p-value = 0.002435
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.615384  -4.324616
## sample estimates:
## mean of the differences
##                   -9.47
```

As the 95% confidence interval does not contain 0, we conclude that the two dosages of Orange juice (0.5 mg and 1 mg) differ in terms of how they affect tooth growth.

```
# Compare the two dosages 1 mg and 2 mg for Orange juice using a paired t-test
supp_2_high <- subset(ToothGrowth, ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2)
t.test(supp_2_mid$len, supp_2_high$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp_2_mid$len and supp_2_high$len
## t = -1.9435, df = 9, p-value = 0.08384
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.2709376  0.5509376
## sample estimates:
## mean of the differences
##                   -3.36
```

Surprisingly, for the first time, our 95% confidence interval contains 0. Also, our p-value is greater than 0.05. Therefore, we fail to reject the null hypothesis and conclude that the dosages 1 mg and 2 mg of Orange juice do not lead to substantial difference in tooth growth.

```r
# Compare the two dosages 0.5 mg and 2 mg for Orange juice using a paired t-test
t.test(supp_2_low$len, supp_2_high$len, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  supp_2_low$len and supp_2_high$len
## t = -7.4919, df = 9, p-value = 3.724e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.703958  -8.956042
## sample estimates:
## mean of the differences
##                  -12.83
```

As was the case with Ascorbic acid, with a very small p-value, we reject the null hypothesis. We conclude that the two dosages of Orange juice (0.5 mg and 2 mg) very clearly differ in terms of how they affect tooth growth.

## Assumptions and conclusions

- Every single t-test was carried out by considering the groups to be paired.
- As the groups were assumed to be paired, we obviously assume equal variances for the groups.
- We assume that the observations are independent and that they are drawn from the same population distribution.
- We conclude that for each supplement, the tooth length generally increases with increase in dosage of the supplement.
- We conclude that the two supplements differ in terms of how they affect tooth growth. We also conclude that Orange juice is the better supplement of the two as the mean tooth length is higher in the case of Orange juice.
- Surprisingly, we conclude that the dosages 1 mg and 2 mg of Orange juice do not lead to substantial difference in tooth growth.