The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Tanguy Levent

26 février 2017

Synopsis

I download the ToothGrowth data in the R datasets package so that analyze the effect of the vitamin C on the Tooth growth in Guinea pigs. First I'm going to provide basic summarises to understand clearly the variables and observations of the dataset. Then I perform some basic exploratory data analyses to finally use confidence intervals and/or hypothesis tests.

Before to go forward, please to take a look on the ToothGrowth R package description

1. Load the ToothGrowth data

I load my dataset and store it in a variable called *"mdata"*.

```
library(dplyr)
library(ggplot2)
library(RColorBrewer)

data("ToothGrowth")
mdata <- ToothGrowth
sum(is.na(mdata))
```

```
## [1] 0
```

Note that there are no missing values in our dataset.

2. Provide a basic summary of the data

```
head(mdata)
```

```
##    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
```

The dataset is composed of the 3 variables : *len* , *supp* and *dose*

```
str(mdata)
```

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

The dataset contains 60 observations. The variable *supp* which represents the delivery methods, have 2 levels inside : "OJ" = "Orange Juice" and "CV" = "Ascorbic Acid".

3. Perform some basic exploratory data analyses

We have understood that the measures taken in the dataset represents tooth growth compared to the dose of vitamin C and the delivery methods. To understand the split by subject/observation I compute a table between *dose* and *supp* :

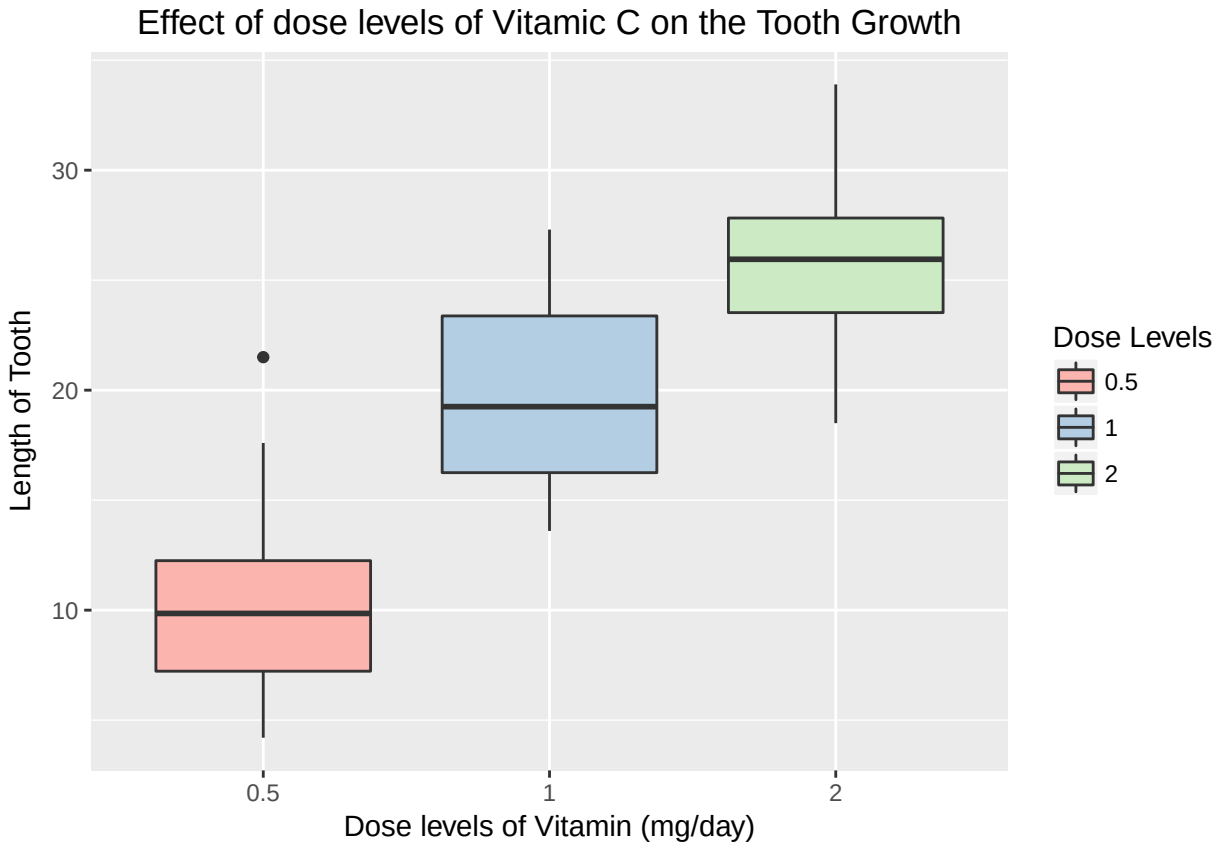
```
table(mdata$dose, mdata$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

In this way, we can observed that the dataset splits 30 pigs by delivery methods and as we have 3 dose levels for each observation, there are 10 subjects by dose levels.

For example we have 10 pigs which had 0.5 mg/day of vitamin C given by Orange Juice method

Now we're going to go more deeper into the analyzes. It will be interesting to see if the dose levels of vitamin C have an impact on the tooth growth. For that I realize with ggplot function a boxplot to vizualise the study (code in Appendix):



Write the mean values by dose levels :

```
## # A tibble: 3 × 2
##    dose    len
##   <dbl>  <dbl>
## 1   0.5 10.605
## 2   1.0 19.735
## 3   2.0 26.100
```

Quickly, thank to the boxplot and the table above, we can conclude that indeed as more the pigs have a daily dose of vitamin C, as more their tooth growth is important.

Now we want to see if the delivery methods of vitamin C as well have an impact on the tooth growth of the pigs. I take the same proceed to realize it :

- Means in terms of delivery methods

```
## # A tibble: 2 × 2
##     supp      len
##   <fctr>    <dbl>
## 1     OJ 20.66333
## 2     VC 16.96333
```

We can conclude that effectively the delivery methods influence the lenght of tooth at each degree of dosage. With the table above we understand that feed pigs of vitamin C with Orange Juice will produce better tooth growth than with Ascorbic Acid.

- Means of Orange Juice by dose

```
## Source: local data frame [3 x 3]
## Groups: supp [1]
##
##     supp  dose   len
##   <fctr> <dbl> <dbl>
## 1     OJ   0.5 13.23
## 2     OJ   1.0 22.70
## 3     OJ   2.0 26.06
```

- Means of Ascorbic Acid by dose

```
## Source: local data frame [3 x 3]
## Groups: supp [1]
##
##     supp  dose   len
##   <fctr> <dbl> <dbl>
## 1     VC   0.5  7.98
## 2     VC   1.0 16.77
## 3     VC   2.0 26.14
```

However, note that at 2 mg/day, there is no difference in the tooth growth between OJ or VC absorption (26.06 VS 26.14).

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

- Hypothesis 1 : mean Orange juice = mean absorbic acid in the tooth growth

```
t.test(len~supp, paired = FALSE, var.equal = FALSE, mdata)
```

```
##
##  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
```

With a p-value equals to 6% slightly above than the threshold of 5% and confidence intervals which include 0 then the null hypothesis cannot be rejected.

As above we can not reject the NULL hypothesis, we go in further details by dose if now we can reject the NULL hypothesis.

- Hypothesis 1.1 : mean Orange juice = mean absorbic acid in the tooth growth with dose = 0.5 mg/day

```
t.test(len~supp, paired = FALSE, var.equal = FALSE, subset(mdata, 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
```

With a p-value equals to 0.6% well below than the threshold of 5% and confidence intervals which not include 0 then the null hypothesis can be rejected. Thus orange juice mean is above of the absorbic mean with a daily dose of 0.5 mg/day (as we expected).

- Hypothesis 1.2 : mean Orange juice = mean absorbic acid in the tooth growth with dose = 1 mg/day

```
t.test(len~supp, paired = FALSE, var.equal = FALSE, subset(mdata, 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
```

With a p-value equals to 0.1% well below than the threshold of 5% and confidence intervals which not include 0 then the null hypothesis can be rejected. Thus orange juice mean is above of the absorbic mean with a daily dose of 1 mg/day (as we expected).

- Hypothesis 1.3 : mean Orange juice = mean absorbic acid in the tooth growth with dose = 2 mg/day

```
t.test(len~supp, paired = FALSE, var.equal = FALSE, subset(mdata, 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
```

With a p-value equals to 96% well above than the threshold of 5% and confidence intervals which include 0 then the null hypothesis cannot be rejected.

- Hypothesis 1.4 : mean Orange juice = mean absorbic acid in the tooth growth without dose = 2 mg/day

```
t.test(len~supp, paired = FALSE, var.equal = FALSE, subset(mdata, dose != 2))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##           17.965           12.375
```

With a p-value equals to 0,4% well below than the threshold of 5% and confidence intervals which not include 0 then the null hypothesis can be rejected.

With this hypothesis we can understand why the null hypothesis 1 can not be rejected, the fault of the dose of 2 mg/day in the data set.

Now we are going to verify like the previous analysis, if our plot analyzes are corrects. Here we want to compare the mean by dose without take in account the delivery methods.

First of all I want to define the correlation function between *dose* and *len*.

```
cor(mdata$dose,mdata$len)
```

```
## [1] 0.8026913
```

The two variables have 80% of correlation which is well positive. Even if 60 observations in the sample is correct to conclude of the relationship between the variables, it's not enough and need to be verify with a test hypothesis.

- Hypothesis 2.1 : mean 0.5 mg/day = mean 1 mg/day in the tooth growth

```
dose1 <- subset(mdata, dose %in% c(0.5,1))
t.test(len~dose, paired = FALSE, var.equal = FALSE, dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

With a p-value equals to 1.268e-07 widely well below of 0.5 and a confidence interval without include 0, the NULL hypothesis that the mean of dose 0.5 and 1 mg/day are equals can be rejected. Mean of 1 mg/day is well above of mean of 0.5 mg/day.

- Hypothesis 2.2 : mean 1 mg/day = mean 2 mg/day in the tooth growth

As we are certificate over 0.5 vs 1 mg/day, we are going to do it over 1 vs 2 mg/day. It's not necessary to do 0.5 vs 2 mg/day as the hypothesis 2.1 have already give us an analysis until 1 mg/day.

```
dose2 <- subset(mdata, dose %in% c(1,2))
t.test(len~dose, paired = FALSE, var.equal = FALSE, dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

As we expected, the p-value is very low, well well below of 5% and there is no 0 in the confidence interval, thus the NULL hypothesis is rejected.

Conclusion

At the first step we have used the exploratory analysis thanks to plots to understand the data set and the variables attached. The first plot shows us an effect of the tooth growth between the dose levels of vitamin C (0.5, 1, 2 mg/day). More the level is important, more the tooth grow. The second plot shows us the impact of the delivery methods of the vitamin C on the tooth growth. We have conclude that Orange Juice gives better results than Ascorbic Acid method delivery. Even if we have seen a decrease impact over the dose level on the Orange Juice because at 2 mg/day, VC is going back at the same mean lenght of tooth than OJ. It could be interesting to add 2 more level doses above 2 mg/day to see if VC get the upper hand over OJ. But exploratory analysis need to be support with statistical analysis so I made it on the same area and its leads us to conclude that the dose of vitamin C and the delivery methods have clearly an impact over the tooth growth.