

The Effect of Vitamin C on Tooth Growth in Guinea Pigs - interpretation of R dataset ToothGrowth

Nikolaos Perdikis

18 7 2019

Summary

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Usage

ToothGrowth

Format

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length

[,2] supp factor Supplement type (VC or OJ).

[,3] dose numeric Dose in milligrams/day

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
data(ToothGrowth)
df_TD <- ToothGrowth
#Display the summary of the dataframe
summary(df_TD)
```

```
##      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
```

```
#Display the unique values of dose
unique(df_TD$dose)
```

```
## [1] 0.5 1.0 2.0
```

Provide a basic summary of the data

```
df_TD$dose <- as.factor(df_TD$dose)
table(df_TD$supp,df_TD$dose)
```

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

Basic stats, so mean:

```
mean(df_TD$len)
```

```
## [1] 18.81333
```

and standard deviation:

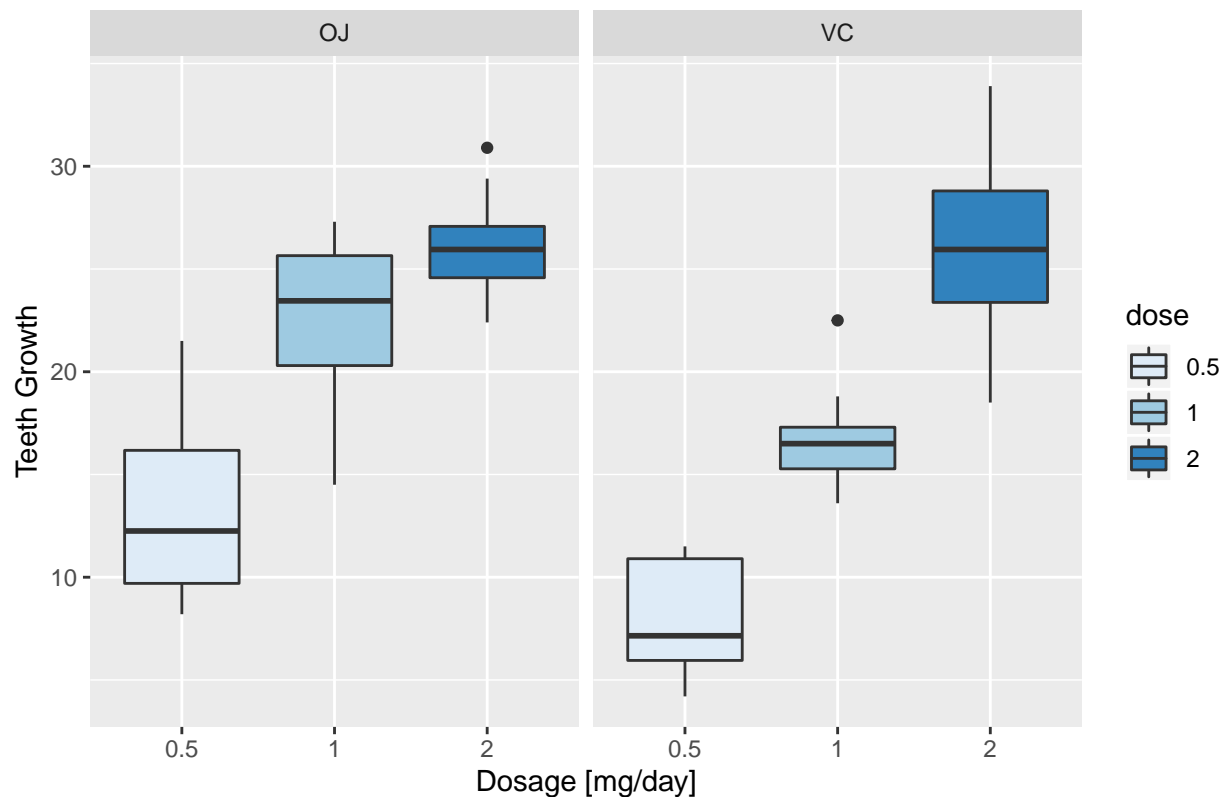
```
sd(df_TD$len)
```

```
## [1] 7.649315
```

What does the table above tell us? We confirm the description of the dataset, as our dataframe contains a total of 60 observations, of 3 different dosage levels and two different supplement types. Let's see how this all looks in a graph:

```
plot <- ggplot(df_TD,
               aes(x=dose,y=len,fill=dose))
plot + geom_boxplot(notch=FALSE) + facet_grid(.~supp) +
  scale_x_discrete("Dosage [mg/day]") +
  scale_y_continuous("Teeth Growth") +
  ggtitle("Effect of Dosage and Supplement Type") +
  scale_fill_brewer(palette="Blues")
```

Effect of Dosage and Supplement Type



While some elements and relationships are clearly visible in the graphic, let's create some filters and summaries:

Get the mean length of toothgrowth, as a function of dose quantity and type of supplementation:

```
TG_1 <- df_TD %>%
  group_by(supp, dose) %>%
  summarize(mean=mean(len), stdev=sd(len), count = n())
print(TG_1)
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp dose  mean stdev count
##   <fct> <fct> <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
```

What about the mean tooth length only as a factor of the supplement?

```
TG_2 <- df_TD %>%
  group_by(supp) %>%
  summarize(mean=mean(len), stdev=sd(len), count = n())
print(TG_2)
```

```
## # A tibble: 2 x 4
##   supp  mean stdev count
```

```
##    <fct> <dbl> <dbl> <int>
## 1 OJ      20.7  6.61    30
## 2 VC      17.0  8.27    30
```

... or the dose?

```
TG_3 <- df_TD %>%
  group_by(dose) %>%
  summarize(mean=mean(len), stdev=sd(len), count = n())
print(TG_3)
```

```
## # A tibble: 3 x 4
##   dose   mean stdev count
##   <fct> <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
```

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

For all tests below, we will assume a 95% confidence interval.

First, let's perform a Student's T-Test comparing the tooth length with the supplement:

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

```
##
## 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
```

What does this tell us?

In plain terms: The type of supplementation did not matter for the tooth length increase, just the dose ... or in statistical phrasing: While comparing a NULL hypothesis (difference of means = 0) against an alternative hypothesis, ** we fail to reject the NULL hypothesis**, since the NULL hypothesis value (delta means = 0) is within the confidence interval of 95% confidence

It would be useful however, to drill into the data, and see how different levels of dosage, in different supplementation type, might affect tooth growth. For this, let's create the relevant data structures:

```
#Reload the data to use comparison signs for dosage:
df_TD <- ToothGrowth
#simple subsetting of the dose level to create new datasets
mindose <- df_TD[df_TD$dose==0.5, ]
meddose <- df_TD[df_TD$dose==1, ]
maxdose <- df_TD[df_TD$dose==2, ]
#High dose ( 2mg/day) and low dose (0.5-1 mg/day) with supplement type OJ: Orange Juice
OJlmdose <- filter(ToothGrowth,dose %in% c(0.5,1),supp=="OJ")
OJmhdose <- filter(ToothGrowth,dose %in% c(1,2),supp=="OJ")
#High dose ( 2mg/day) and low dose (0.5-1 mg/day) with supplement type VC: Ascorbic Acid
```

```
VClddose <- filter(ToothGrowth,dose <2,supp=="VC")
VCmhdose <- filter(ToothGrowth,dose > 0.5 ,supp=="VC")
```

Let's compare, at the **0.5 - 1 mg/day dosage levels** what happens between Orange Juice and Ascorbic Acid

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

```
##
## 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
```

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

```
##
## 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
```

Conclusion: We **reject the NULL hypothesis** or in simple terms, different supplementation method for 0.5 - 1 mg/day **does yield different results** in tooth growth.

Let's compare, at the **2 mg/day dosage levels** what happens between Orange Juice and Ascorbic Acid

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

```
##
## 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
```

Conclusion: We **fail reject the NULL hypothesis** or in simple terms, it makes no difference if our guinea pigs receive the Vitamin C in Orange Juice or otherwise.

Since we are on it, let's also apply Student's T-Test in the rest of our datasets. So we will investigate if: We can separate between **0.5 and 1 mg/day**, for supplementation type to be **Orange Juice**:

```
t.test(len ~ dose, paired=FALSE, var.equal=FALSE, data=OJlmdose)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -13.415634 -5.524366
## sample estimates:
## mean in group 0.5 mean in group 1
## 13.23 22.70
```

This test shows us that there is a significant difference between the two dosages of 0.5 and 1 mg/day, when supplied with Orange Juice (OJ)

We can separate between **1 and 2 mg/day**, for supplementation type to be **Orange Juice**:

```
t.test(len ~ dose, paired=FALSE, var.equal=FALSE, data=OJmhdose)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.5314425 -0.1885575
## sample estimates:
## mean in group 1 mean in group 2
## 22.70 26.06
```

This test shows us that there is a significant difference.

We can separate between **0.5 and 1 mg/day**, for supplementation type to be **Ascorbic Acid**:

```
t.test(len ~ dose, paired=FALSE, var.equal=FALSE, data=VClmdose)
```

```
##
## Welch Two Sample t-test
##
## data: len by dose
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.265712 -6.314288
## sample estimates:
## mean in group 0.5 mean in group 1
## 7.98 16.77
```

This test shows us that there is a significant difference.

We can separate between **1 and 2 mg/day**, for supplementation type to be **Ascorbic Acid**:

```
t.test(len ~ dose, paired=FALSE, var.equal=FALSE, data=VCmhdose)
```

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

```
##
## data:  len by dose
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean in group 1 mean in group 2
##           16.77           26.14
```

This test shows us that there is a significant difference.

State your conclusions and the assumptions needed for your conclusions

Conclusions:

- The amount of tooth length increase is directly analogous to the vitamin intake, regardless of type of injection
- Given a cumulative grouping, the two vitamin injection types yield similar results, with a 95% confidence interval
- Given regard to the amount of the dose, the type of injection of the vitamin matters for low to mid dose, but does not for the max dose of 2mg / day

The following assumptions were made:

- The measurements are not paired
- We do not assume the variances to be equal
- We assume that our population samples are IID