The Effect of Vitamin C on Teeth Growth: A T-Student Statistical Inference

Francisco Ganhão

10/20/2016

Synopsis

The present document analyzes the *ToothGrowth* dataset from *R*'s *datasets* package, through exploratory analysis, basic statistics, and hypothesis testing.

The *ToothGrowth* dataset studies the tooth growth of 60 guinea pigs that were supplemented with two types of vitamin C supplements, orange juice (*OJ*) and ascorbic acid (*VC*), and for each of them three dose amounts: 0.5 mg/day; 1 mg/day; and 2 mg/day.

The document poses the following two hypothesis: 1) if increasing the uptake of vitamin C increases the tooth growth; 2) if orange juice yields a higher tooth growth than ascorbic acid.

At the end of the document some remarks will be added concerning the hypothesis testing of the data.

Exploratory Analysis

Before delving through the *ToothGrowth* dataset, let us load the datasets package.

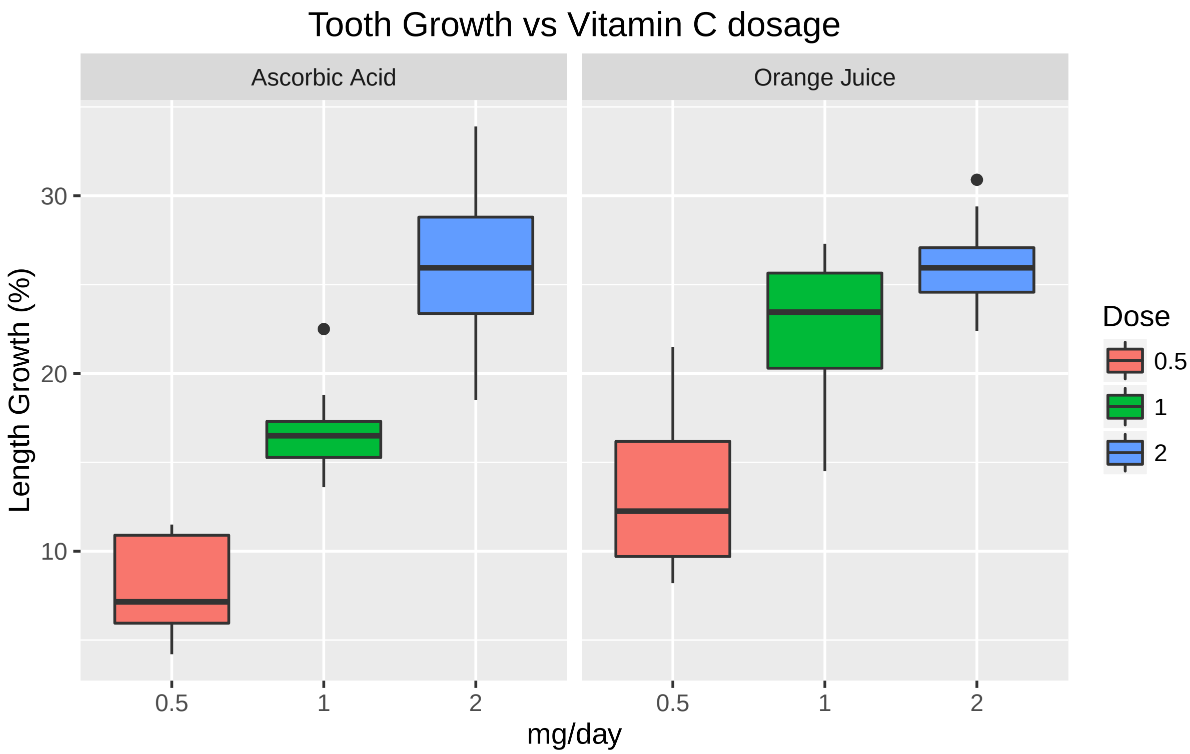
```
library(datasets,quietly = TRUE)
```

For a clearer understanding of the exploratory analysis, let us modify the dataset using the *dplyr* package.

```
library(dplyr, quietly = TRUE, warn.conflicts = FALSE)
#Alter the column names
dataset <- ToothGrowth %>%
            select(Length = len, Type = supp, Dose = dose) %>%
            mutate(Type = as.character(Type))

#Alter OJ to Orange Juice and VC as Vitamin C
dataset[dataset[,"Type"] == "OJ","Type"] <- "Orange Juice"
dataset[dataset[,"Type"] == "VC","Type"] <- "Ascorbic Acid"
dataset <- dataset %>%
            mutate(Type = as.factor(Type), Dose = as.factor(Dose))
```

Using the ggplot package, let us plot tooth growth in function of the doses and respective supplement type and for the sake of simplicity let's assume that the tooth growth is in percentage (%).



As observed from the data both orange juice and the ascorbic acid, seem to increase the tooth growth as the dosage of vitamin C increases.

At a first glance, the orange juice seems to yield a better tooth growth when compared to the ascorbic acid, although they look almost equivalent for 2 $mg/$day of vitamin C.

Data Summary

In this section, the reader can find basic statistics that summarize the data for the amount of vitamin C provided to the Guinea Pigs in general; and both the amount and supplement type.

Each sub-section will focus on the aforementioned topics where the results are displayed in tables in function of the dose, where first it's analyzed for both supplements and then for each supplement separately.

Each sub-section presents a table with basic statistics in function of the vitamin C dose (presented as Dose in the tables most left column).

The analyzed statistics for each table are the following: average growth (Average column); growth standard deviation (Deviation column); minimum growth (Minimum column); maximum growth (Maximum column); growth first quantile (Q1 column); growth median (Median column); third quantile (Q3 column); and the number of samples (Samples column).

As a remark there is a total of 60 samples in this study.

Growth based on Vitamin C Dose

```
library(knitr, quietly = TRUE)
table <-
    group_by(dataset,Dose) %>%
    summarise(Average = mean(Length),
```

```
              Deviation = sd(Length),
              Min = quantile(Length,0),
              Q1 = quantile(Length,0.25),
              Median = quantile(Length,0.5),
              Q3 = quantile(Length,0.75),
              Max = quantile(Length,1),
              Samples = length(Length))
kable(table)
```

Dose	Average	Deviation	Min	Q1	Median	Q3	Max	Samples
0.5	10.605	4.499763	4.2	7.225	9.85	12.250	21.5	20
1	19.735	4.415436	13.6	16.250	19.25	23.375	27.3	20
2	26.100	3.774150	18.5	23.525	25.95	27.825	33.9	20

Growth based on Orange Juice

```
table <-
    filter(dataset, Type == "Orange Juice") %>%
    group_by(Dose) %>%
    summarise(Average = mean(Length),
              Deviation = sd(Length),
              Min = quantile(Length,0),
              Q1 = quantile(Length,0.25),
              Median = quantile(Length,0.5),
              Q3 = quantile(Length,0.75),
              Max = quantile(Length,1),
              Samples = length(Length))
kable(table)
```

Dose	Average	Deviation	Min	Q1	Median	Q3	Max	Samples
0.5	13.23	4.459708	8.2	9.700	12.25	16.175	21.5	10
1	22.70	3.910953	14.5	20.300	23.45	25.650	27.3	10
2	26.06	2.655058	22.4	24.575	25.95	27.075	30.9	10

Growth based on Ascorbic Acid

```
table <-
    filter(dataset, Type == "Ascorbic Acid") %>%
    group_by(Dose) %>%
    summarise(Average = mean(Length),
              Deviation = sd(Length),
              Min = quantile(Length,0),
              Q1 = quantile(Length,0.25),
              Median = quantile(Length,0.5),
              Q3 = quantile(Length,0.75),
              Max = quantile(Length,1),
```

```
              Samples = length(Length))
kable(table)
```

Dose	Average	Deviation	Min	Q1	Median	Q3	Max	Samples
0.5	7.98	2.746634	4.2	5.950	7.15	10.9	11.5	10
1	16.77	2.515309	13.6	15.275	16.50	17.3	22.5	10
2	26.14	4.797731	18.5	23.375	25.95	28.8	33.9	10

Confidence Intervals and Hypothesis Tests

In the present section, the following hypotheses will be tested:

- **A. Does increasing the vitamin C dose implies a higher tooth growth?**
- **B. Does orange juice have a higher benefit on tooth growth than ascorbic acid under the same vitamin C dose?**

Hypothesis **A** studies the teeth growth by **dose**, while hypothesis **B** studies the teeth growth by **supplement**.

Since for both hypotheses we are dealing with a small number of samples, the *t-test* distribution should be used to test our assumptions.

In addition to the previous and for the sake of simplicity let us define a function that based on two independent groups and a given α, as our Type I error, computes the following items for a t-distribution confidence interval: 1) lower limit of the confidence interval; 2) upper limit of the confidence interval; 3) difference between the means of the groups; 4) the degrees of freedom; 5) the adjusted deviation of the t-test distribution.

```
confidence <- function(group1, group2, alpha){
    n1 <- length(group1)
    n2 <- length(group2)
    u1 <- mean(group1)
    u2 <- mean(group2)
    v1 <- var(group1)
    v2 <- var(group2)
    df <- (v1/n1 + v2/n2)^2 /
          ((v1/n1)^2/(n1-1) + (v2/n2)^2/(n2-1))
    s <- sqrt(v1/n1+v2/n2)
    interval <- mean(group1) - mean(group2) + c(-1,1)*qt(1-alpha/2, df)*s
    c(interval[1], interval[2], mean(group1) - mean(group2), df, s)
}
```

A. T-Test: Does Increasing the Vitamin C Dose Implies a Higher Tooth Growth?

A.1. Tooth Growth Between 0.5 mg/day and 1 mg/day of Supplements

We would like to confirm if using 1 mg/day of vitamin C yields a higher tooth growth than 0.5 mg/day.

Below we have our null hypothesis, H_0, and its the respective alternative, H_a, where the average length growth for 1 mg is defined as $\mu_{1.0}$ and for 0.5 mg as $\mu_{0.5}$.

- H_0: $\mu_{1.0} = \mu_{0.5}$, **where the average growth is the same.**
- H_a: $\mu_{1.0} > \mu_{0.5}$, **where the average growth is higher for a greater dose.**

Assuming a **one-sided** confidence interval of 95%, yields the following confidence interval for the difference between the two groups, where *tdf* stands for degrees of freedom and *p-value* for the samples' p-value.

```
g05 <- filter(dataset, Dose == "0.5")$Length
g1 <- filter(dataset, Dose == "1")$Length
interval <- confidence(g05, g1, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(Lower = -Inf, Center = interval[3], Upper = interval[2],
                    tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	-9.13	-6.753323	37.98641	1e-07

Since we wish to test that $\mu_{1.0} = \mu_{0.5}$, we can say that for a confidence interval of 95%, we can **reject the null hypothesis** since the p-value is lower than 5%.

A.2. Tooth Growth Between 0.5 mg/day and 2 mg/day of Supplements

As in the previous section we would like to confirm if 2 mg/day of vitamin C has the same benefits as 0.5 mg/day.

Below, we have the null hypothesis H_0, and its respective alternative H_a, where the the average length growth for 2 mg is $\mu_{2.0}$ and for 0.5 mg is $\mu_{0.5}$.

- H_0: $\mu_{2.0} = \mu_{0.5}$, **where the average growth is the same**.
- H_a: $\mu_{2.0} > \mu_{0.5}$, **where the average growth is higher for a greater dose.**

Below we have the parameters for a **one-sided** confidence interval of 95%.

```
g2 <- filter(dataset, Dose == "2")$Length
interval <- confidence(g05, g2, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(
            Lower = -Inf, Center = interval[3], Upper = interval[2],
            tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	-15.495	-13.27926	36.88258	0

Since we wish to test that $\mu_{2.0} = \mu_{0.5}$, we can say that for a confidence interval of 95%, we can **reject the null hypothesis** since the p-value is less than 5%.

A.3. Tooth Growth Between 1 mg/day and 2 mg/day of Supplements

Now we wish to compare the null hypothesis H_0, where both 1 mg and 2 mg have the same benefit, or the alternative, H_a, where a higher dose has a higher benefit.

- **H_0: $\mu_{2.0} = \mu_{1.0}$, where the average growth is the same.**
- **H_a: $\mu_{2.0} > \mu_{1.0}$, where the average growth is higher for a greater dose.**

Below we have the parameters for a **one-sided** confidence interval of 95%.

```
interval <- confidence(g1, g2, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(
            Lower = -Inf, Center = interval[3], Upper = interval[2],
            tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	-6.365	-4.17387	37.10109	9.5e-06

Since we wish to test that $\mu_{2.0} = \mu_1$, we can say that for a confidence interval of 95%, we can **reject the null hypothesis** since the p-value shown above is lower than 5%.

A.4. Supplement Dose Remarks

As seen in subsections *A.1, A.2* and *A.3*, we can definitely conclude that increasing the supplement dose does imply a higher tooth growth, since the null-hypothesis of the tooth growth being the same for all doses has been rejected.

B. Does Orange Juice Have a Higher Benefit on Tooth Growth than Ascorbic Acid Under the Same Vitamin C Dose?

B.1. Tooth Growth Between Ascorbic Acid and Orange Juice for 0.5 mg/day

We would like to confirm if using 0.5 mg/day of vitamin C yields a higher tooth growth for orange juice than it does for ascorbic acid.

Below we have our null hypothesis, H_0, and its the respective alternative, H_a, where we have the average length growth for ascorbic acid mg, μ_{VC} and μ_{OJ} for orange juice.

- **H_0: $\mu_{OJ} = \mu_{VC}$, where the average growth is the same.**
- **H_a: $\mu_{OJ} > \mu_{VC}$, where the average growth is higher for orange juice.**

Assuming a **one-sided** confidence interval of 95%, yields the following confidence interval for the difference between the two groups, where *tdf* stands for degrees of freedom and *p-value* for the samples' p-value.

```
gOJ <- filter(dataset, Dose == "0.5" & Type == "Orange Juice")$Length
gVC <- filter(dataset, Dose == "0.5" & Type == "Ascorbic Acid")$Length
interval <- confidence(gVC, gOJ, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(Lower = -Inf, Center = interval[3], Upper = interval[2],
                    tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	-5.25	-2.34604	14.96875	0.0031793

Since we wish to test that $\mu_{OJ} = \mu_{VC}$, we can say that for a confidence interval of 95%, we can **reject the null hypothesis** since a p-value of 0.3% is lower than 5%.

B.2. Tooth Growth Between Ascorbic Acid and Orange Juice for 1 mg/day

We would like to confirm if using 1 mg/day of vitamin C yields a higher tooth growth for orange juice than it does for ascorbic acid - below we have the null hypothesis, H_0 and the respective alternative, H_a.

- **H_0: $\mu_{OJ} = \mu_{VC}$, where the average growth is the same.**
- **H_a: $\mu_{OJ} > \mu_{VC}$, where the average growth is higher for orange juice.**

Assuming a **one-sided** confidence interval of 95%, yields the following confidence interval for the difference between the two groups, where *tdf* stands for degrees of freedom and *p-value* for the samples' p-value.

```r
gOJ <- filter(dataset, Dose == "1" & Type == "Orange Juice")$Length
gVC <- filter(dataset, Dose == "1" & Type == "Ascorbic Acid")$Length
interval <- confidence(gVC, gOJ, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(Lower = -Inf, Center = interval[3], Upper = interval[2],
                    tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	-5.93	-3.356158	15.35767	0.0005192

Since we wish to test that $\mu_{OJ} = \mu_{VC}$, we can say that for a confidence interval of 95%, we can **reject the null hypothesis** since a p-value of 0.05% is lower than our 5% interval.

B.3. Tooth Growth Between Ascorbic Acid and Orange Juice for 2 mg/day

We would like to confirm if using 2 mg/day of vitamin C yields a higher tooth growth for orange juice than it does for ascorbic acid - below we have the null hypothesis, H_0 and the respective alternative, H_a.

- **H_0: $\mu_{OJ} = \mu_{VC}$, where the average length growth is the same.**
- **H_a: $\mu_{OJ} > \mu_{VC}$, where the average growth is higher for orange juice.**

Assuming a **one-sided** confidence interval of 95%, yields the following confidence interval for the difference between the two groups, where *tdf* stands for degrees of freedom and *p-value* for the samples' p-value.

```r
gOJ <- filter(dataset, Dose == "2" & Type == "Orange Juice")$Length
gVC <- filter(dataset, Dose == "2" & Type == "Ascorbic Acid")$Length
interval <- confidence(gVC, gOJ, 0.1)
pvalue <- pt(interval[3]/interval[5], df = interval[4])
table <- data.frame(Lower = -Inf, Center = interval[3], Upper = interval[2],
```

```
                    tdf = interval[4], pvalue = pvalue)
names(table)[5] <- "p-value"
kable(table)
```

Lower	Center	Upper	tdf	p-value
-Inf	0.08	3.1335	14.03982	0.5180742

Since we wish to test that $\mu_{OJ} = \mu_{VC}$, we can say that for a confidence interval of 95%, we **fail to reject the null hypothesis** since our p-value is higher than our 5% interval.

B.4. Supplement Type Remarks

Based on the sub-sections *B.1, B.2* and *B.3*, we can say that the orange juice improves the tooth growth compared to the ascorbic acid, for supplement doses of 0.5 mg and 1 mg of vitamin C; though for a dose of 2 mg, the effect of the orange juice is similar to the ascorbic acid, since we have failed to reject the null hypothesis.

Conclusions

Based on the aforementioned results, we have observed that increasing the dose of vitamin C is instrumental for the guinea pigs' teeth growth since statistic results show that these are significant.

On the other hand it was observed that for doses, other than 2 mg of vitamin C, orange juice has a greater effect on the teeth growth than using ascorbic acid since statistical results were significant enough; however for 2 mg, statistical results were not significant enough to conclude that there's a better benefit to use orange juice than ascorbic acid.

Despite these conclusions, it would be interesting to see the effect of other vitamin C quantities.