# Effectiveness of Vitamin C supplements on teeth grow of guinea pigs

*Ion Scerbatiuc*

*September 26, 2015*

## Synopsis

In the following analysis we are trying to analyze the effect of two supplements, each at three different doses on the tooth length of guinea pigs. We will first explore the data and come up with a hypothesis about the effectiveness of one suppliment over the other. Next we will use hypothesis testing to show that our hypothesis is correct.

```
library(reshape2)
library(ggplot2)
library(datasets)
library(dplyr)
library(xtable)
options(xtable.comment = FALSE)
```

## Exploring the data

Let's first load the data and see what the variables are and how the data is organized.

```
data(ToothGrowth)
print(
    xtable(head(ToothGrowth), caption="ToothGrowth dataset preview"),
    include.rownames=FALSE,
    sanitize.text=identity)
```

| len | supp | dose |
|---|---|---|
| 4.20 | VC | 0.50 |
| 11.50 | VC | 0.50 |
| 7.30 | VC | 0.50 |
| 5.80 | VC | 0.50 |
| 6.40 | VC | 0.50 |
| 10.00 | VC | 0.50 |

Table 1: ToothGrowth dataset preview

We can see the three variables of interest:

- `len` - The observed tooth lenght
- `supp` - The type of suplement, where `OJ` means Orange Juice and `VC` means Vitamin C
- `dose` - The dose in milligrams

Let's now find the summary statistics of tooth growth by supplement and dose:

```
data_summary <- group_by(ToothGrowth, supp, dose) %>%
    summarise(count = n(), mean = mean(len), median = median(len), stddev = sd(len))
print(
    xtable(data_summary, caption="Tooth growth by supplement and dose"),
    include.rownames=FALSE,
    sanitize.text=identity)
```

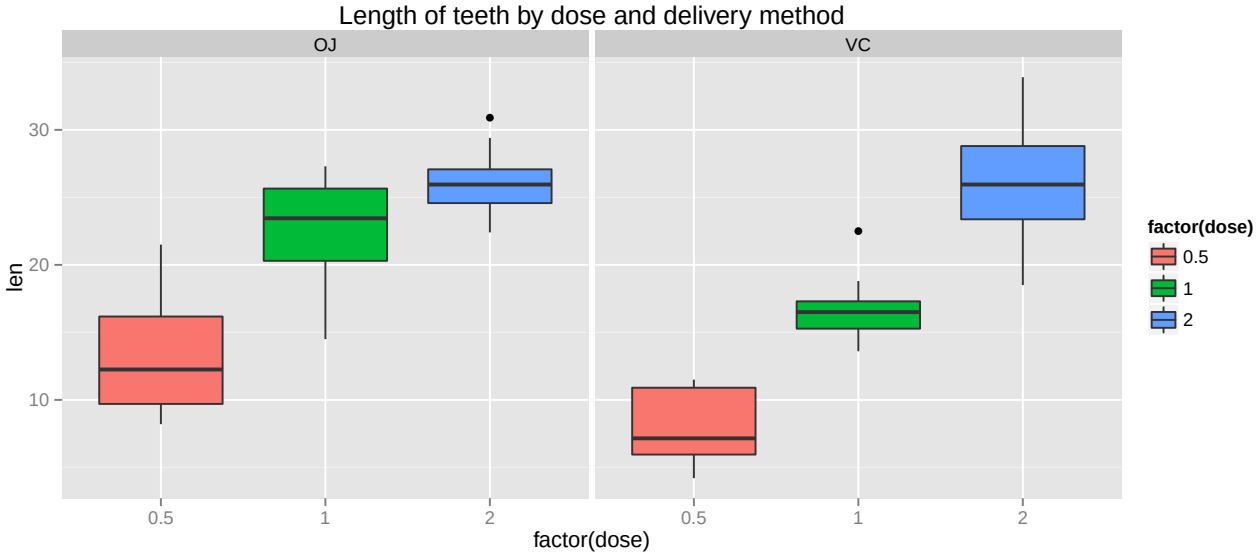
| supp | dose | count | mean | median | stddev |
|---|---|---|---|---|---|
| OJ | 0.50 | 10 | 13.23 | 12.25 | 4.46 |
| OJ | 1.00 | 10 | 22.70 | 23.45 | 3.91 |
| OJ | 2.00 | 10 | 26.06 | 25.95 | 2.66 |
| VC | 0.50 | 10 | 7.98 | 7.15 | 2.75 |
| VC | 1.00 | 10 | 16.77 | 16.50 | 2.52 |
| VC | 2.00 | 10 | 26.14 | 25.95 | 4.80 |

Table 2: Tooth growth by supplement and dose

Looking at the table above we can intuitively guess that the Orange Juice supliment is on average more effective than the Vitamic C one, especially for the smaller doses of 0.5 and 1 mg.

Let's confirm our intuition by plotting the data:

```
ggplot(ToothGrowth, aes(x = factor(dose), y = len, fill=factor(dose))) +
    geom_boxplot() +
    facet_grid(. ~ supp) +
    ggtitle("Length of teeth by dose and delivery method")
```




The plots above shows us that indeed the Orange Juice seems more effective than the other supplemenet.

## Hypothese testing

Lets formalise the above observation in form of a hypothesis: Based on the observed data about a sample of 10 subjects, we believe that the Orange Juice supplement is more effective on the teeth length than the Vitamic C supplement.

First, let's measure the effect of the supplement on tooth growth by performing a two-sample t-test for the difference in tooth length by suplement type.

```
results <- t.test(len ~ supp, data = ToothGrowth, var.equal = TRUE, paired=FALSE)
df_results <- data.frame(
    OJ.mean = results$estimate[1],
    VC.mean = results$estimate[2],
    Conf.Int.Low = results$conf.int[1],
    Conf.Int.High = results$conf.int[2],
    P.value = results$p.value
)
print(
    xtable(df_results, caption="Tooth growth by supplement t-test summary"),
    include.rownames=FALSE,
    sanitize.text=identity)
```

| OJ.mean | VC.mean | Conf.Int.Low | Conf.Int.High | P.value |
|---|---|---|---|---|
| 20.66 | 16.96 | -0.17 | 7.57 | 0.06 |

Table 3: Tooth growth by supplement t-test summary

From above table we extract the following insights:

- the `confidence interval` contains zero, which means that there is a probability that the effects are the same (difference of lengths can be equal to zero)
- the `p-value` is larger than `0.05`, which means that the evidence is not statistically significant for our null hypothesis.

Considering the two above, we can't prove with 95% certainty that the Orange Juice supplement is more effective than the Vitamic C one.

On the other hand the data showed us that there seems to be a difference, at least for the smaller doses. Let's then try to run a set of similar two-sample t-tests for each dose in the dataset.

```
doses = unique(ToothGrowth$dose)
results <- lapply(doses, function (d) {
    filtered_data <- filter(ToothGrowth, dose == d)
    t.test(len ~ supp, data = filtered_data, var.equal = TRUE, paired=FALSE)
})

df_results <- data.frame(
    Dose = doses,
    OJ.mean = sapply(results, function (r) r$estimate[1]),
    VC.mean = sapply(results, function (r) r$estimate[2]),
    Conf.Int.Low = sapply(results, function (r) r$conf.int[1]),
    Conf.Int.High = sapply(results, function (r) r$conf.int[2]),
    P.value = sapply(results, function (r) r$p.value)
)
```

In the above snipped we extract the list of unique doses in the dataset and run, using `lapply` and an anonymous function, the t-test for the two samples for each unique dose.

And finally let's see what the test results are:

```r
print(
    xtable(df_results, caption="Tooth growth by supplement and dose t-test summary"),
    include.rownames=FALSE,
    sanitize.text=identity)
```

| Dose | OJ.mean | VC.mean | Conf.Int.Low | Conf.Int.High | P.value |
|---|---|---|---|---|---|
| 0.50 | 13.23 | 7.98 | 1.77 | 8.73 | 0.01 |
| 1.00 | 22.70 | 16.77 | 2.84 | 9.02 | 0.00 |
| 2.00 | 26.06 | 26.14 | -3.72 | 3.56 | 0.96 |

Table 4: Tooth growth by supplement and dose t-test summary

From above table we extract the following insights:

- For the first two doses (`0.5` and `1 mg` respectively), the t-test yields confidence intervals that are above zero with `p-values` that are lower than `0.05`. That means that there is indeed a statistically significant difference between the average teeth lengths in the two groups
- For the `2 mg` dose, the confindence interval is almost perfectly symmetric around zero and the `p-value` is almost `1` which means that there is almost no difference in the length averages between the two groups.

In conclusion, based on the data we have, we can say with 95% confidence that the Orange Juice supplement is more effective that the Vitamin C for smaller dosages of the respective supplement (`0.5` and `1 mg` respectively). If the dose is increased however (to `2 mg`), the two types of supplements perform almost identically.