

# Title: Impact of different vitamin C dose levels and supplement types on tooth lengths of Guinea pig

Author: midham76

Date: 20 October 2015

## Overview

This experiment will analyze the `ToothGrowth` data of the R data sets package which describes the effect of vitamin C on tooth growth in guinea pigs

The `ToothGrowth` data set consists of 60 observations of 3 variables:

- `len`: Tooth length in millimeters (numeric variable)
- `supp`: Supplement type (factor variable with levels VC and OJ)
- `dose`: Dose in milligrams (numeric variable)

Dose will be treated as a factor in the subsequent analysis. Only three dose levels of Vitamin C (0.5, 1, and 2 mg) were used.

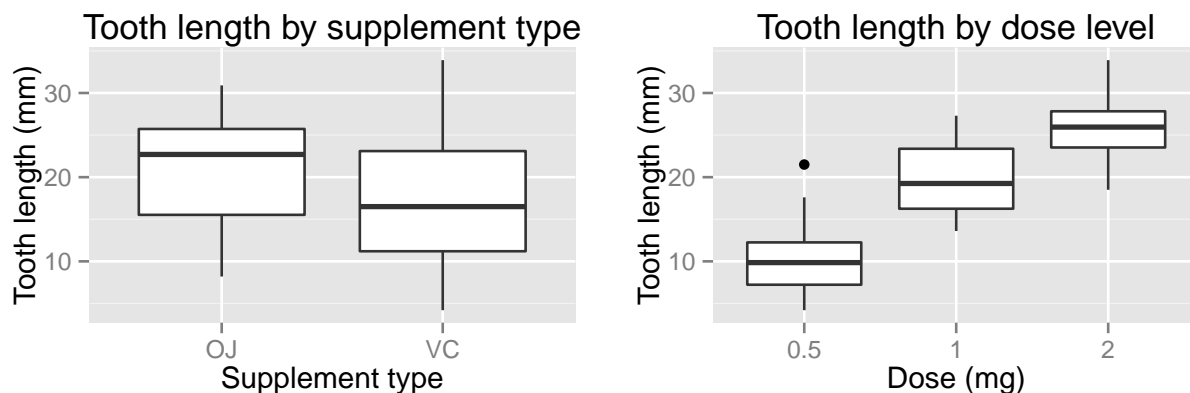
A total of 60 unique guinea pigs will be used. Each guinea pig was assigned to a group which received a specific dose level of vitamin C. In addition, in each of those groups two different delivery methods (orange juice or ascorbic acid) were applied leaving 10 pigs per subgroup:

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

## Exploratory data analysis

The average guinea pig tooth length is 18.813 with a standard deviation of 7.649.

However, the experiment is interested in finding differences in tooth length across different groups of supplement types, dose levels and their respective combinations.

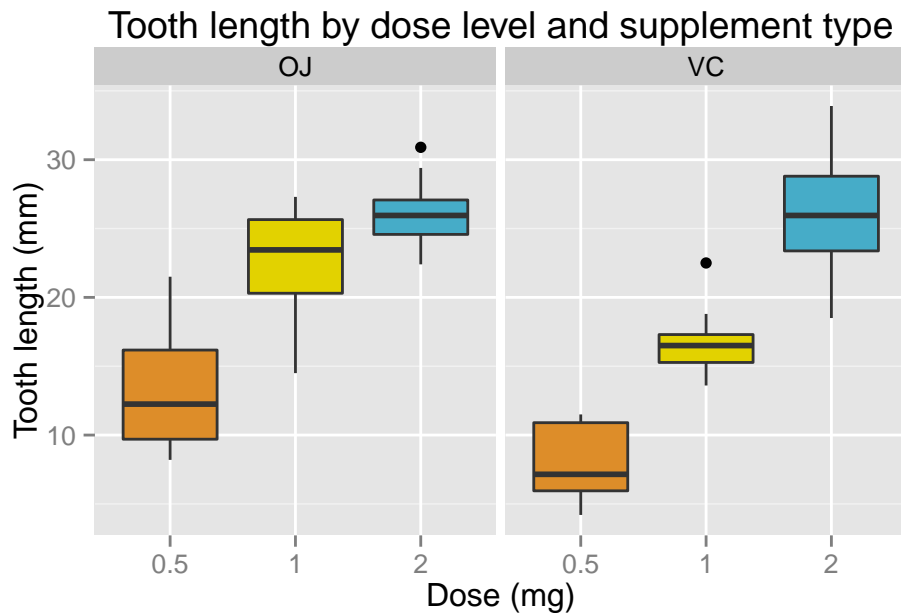


The left box plot above depict that guinea pigs which received their dose of vitamin C using orange juice have on average a greater tooth length than those pigs which got their dose using ascorbic acid. The orange juice group distribution is skewed to the left whereas the ascorbic acid group seems to be fairly symmetric.

Grouping by dose level (right hand side box plot) reveals even greater differences in tooth length and their respective means. The differences between the dose level groups are so big that not even the interquartile range boundaries overlap:

```
##   dose   q25   q75
## 1  0.5  7.225 12.250
## 2  1.0 16.250 23.375
## 3  2.0 23.525 27.825
```

Both trends can be further observed when examining tooth length differences across both groups at the same time.



## Hypothesis tests

### Conditions

As the experiment is dealing with subsets of  $n \leq 30$ , standard error estimate will generally not be accurate. Therefore, experiment will use the t distribution for our hypothesis tests and when constructing confidence intervals.

Based on the current information it is hard to check if the 2 necessary prerequisites for this setup are met:

1. Independence of the observations
2. Observations come from a nearly normal distribution.

However, the experiment assumed that the guinea pigs were randomly assigned to one of the groups (guarantees independence) and that they were sampled from a nearly normal population.

## Differences in supplement types

The experiment will check if the observed difference in tooth length means between pigs who received their dose using orange juice and those who received their dose via ascorbic acid is statistically different from 0.

The conservative choice will be a t-test with unequal variances of the two samples. The table below shows the results:

```
## estimate estimate1 estimate2 statistic p.value parameter conf.low
## 1 3.7 20.66333 16.96333 1.915268 0.06063451 55.30943 -0.1710156
## conf.high
## 1 7.571016
```

With a p-value of 0.061 which is larger than the significance value, 0.05, the experiment failed to reject the null hypothesis. The data does not provide strong evidence that the average difference in tooth length across the two supplement types is different from 0.

The hypothesis test result corresponds to the respective confidence interval -0.171, 7.571 because it includes 0.

## Differences in dose levels

Since the experiment are dealing with 3 different dose levels we need to apply 3 t-tests to cover all factor level combinations. Again, our null hypothesis in all cases is that the average difference in tooth length is 0.

The table below summarizes the results of the three 3 pairwise comparisons. Column **null\_hypothesis** shows which null hypothesis was used for the respective t-test also indicating which sub group dose levels means were compared to each other.

```
## null_hypothesis estimate estimate1 estimate2 statistic p.value
## 1  $\hat{\mu}_{2mg} - \hat{\mu}_{1mg} = 0$  6.365 26.100 19.735 4.900484 1.906430e-05
## 2  $\hat{\mu}_{2mg} - \hat{\mu}_{0.5mg} = 0$  15.495 26.100 10.605 11.799046 4.397525e-14
## 3  $\hat{\mu}_{1mg} - \hat{\mu}_{0.5mg} = 0$  9.130 19.735 10.605 6.476648 1.268301e-07
## parameter conf.low conf.high
## 1 37.10109 3.733519 8.996481
## 2 36.88259 12.833833 18.156167
## 3 37.98641 6.276219 11.983781
```

In all three cases the p-value is smaller than the significance level 0.05 which means we reject the null hypothesis each time. The data provides strong evidence that the average tooth length of guinea pigs is different for each dose level.

Interpreting the first row of the table for illustration:

The experiment is 95% confident that the average tooth length of guinea pigs who received a 2mg dose of vitamin C is on average 3.73 to 9 millimeters higher than those who received a 1mg dose.

## Conclusions

The experiment was able to show that the observed difference of tooth length mean across supplement types is statistically not different from 0.

The experiment also highlight the fact that tooth length indeed varies across groups of dose levels and that all of those differences are statistical and pramatically significant.

A latter study might take into account pairwise t-tests of the different supplement types/dose level combinations.

```

library(ggplot2)
library(dplyr)
library(broom)
library(wesanderson)
library(grid)

# Tables

with(ToothGrowth, table(dose, supp))

ToothGrowth %>% group_by(dose) %>%
  summarize(
    q25 = quantile(len, 0.25),
    q75 = quantile(len, 0.75)
  ) %>%
  as.data.frame

# Generate plots

g_len_by_supp <- ggplot(ToothGrowth, aes(supp, len)) +
  geom_boxplot() +
  xlab('Supplement type') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by supplement type')

g_len_by_dose <- ggplot(ToothGrowth, aes(as.factor(dose), len)) +
  geom_boxplot() +
  xlab('Dose (mg)') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by dose level')

ggplot(ToothGrowth, aes(as.factor(dose), len)) +
  geom_boxplot(aes(fill = as.factor(dose))) +
  facet_grid(. ~ supp) +
  xlab('Dose (mg)') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by dose level and supplement type') +
  scale_fill_manual(values = wes_palette("FantasticFox")) +
  theme(legend.position = "none")

# T-test for difference in tooth length means across supplement types

t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE)
tidy(t_diff_supp)

# Pairwise t-tests for differences in tooth lengths means across dose levels

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
  ToothGrowth$len[ToothGrowth$dose == 1]) %>%

```

```

tidy %>%
mutate(
  null_hypothesis = 'μ2mg - μ1mg = 0'
) %>%
select(9, 1:8)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
  ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
tidy %>%
mutate(
  null_hypothesis = 'μ2mg - μ0.5mg = 0'
) %>%
select(9, 1:8) %>%
bind_rows(pairwise_results, .)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 1],
  ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
tidy %>%
mutate(
  null_hypothesis = 'μ1mg - μ0.5mg = 0'
) %>%
select(9, 1:8) %>%
bind_rows(pairwise_results, .)

print.data.frame(pairwise_results)

```