title: "Impact of different vitamin C dose level and supplement types on tooth length of Guinea pigs" author: "Sumant" date: "29/08/2020"

output: html_document

````{r echo=FALSE, warning=FALSE, message=FALSE} library(ggplot2) library(dplyr) library(broom) library(wesanderson) library(grid)

```
## Overiew

In this short report we are going to analyze the `TootGrowth` data of the R data sets package which describes the effect of vitamin

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)

Even though `dose` is a numeric variable we will treat it as a factor in the subsequent analysis because only three dose levels of V

In total, we deal with 60 unique guinea pigs. Each pig was assigned to a group which received a specific dose level of vitamin C. In

```{r echo=FALSE}
with(ToothGrowth, table(dose, supp))
```

## Exploratory data analysis

The average guinea pig tooth length is `r round(mean(ToothGrowth$len), 3)` with a standard deviation of `r round(sd(ToothGrowth$len), 3)`.

However, we are more interested in finding differences in tooth length across different groups of supplement types, dose levels and their respective combinations.

````{r echo=FALSE, fig.width=7, fig.height=2.5, fig.align='center'} 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') grid.newpage()

# Push viewport to graphic device

pushViewport(viewport(layout = grid.layout(1, 2, widths = c(0.5, 0.5)))) print(g_len_by_supp, vp = viewport(layout.pos.row = 1, layout.pos.col = 1)) print(g_len_by_dose, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))

```
The left box plot above reveals that guinea pigs which received their dose of vitamin C using orange juice have on average a greater

Grouping by dose level (right hand side box plot) reveals even greater differences in tooth length and their respective means. The d

```{r echo = FALSE}
ToothGrowth %>% group_by(dose) %>%
  summarize(
    q25 = quantile(len, 0.25),
    q75 = quantile(len, 0.75)
  ) %>%
  as.data.frame
```

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

````{r echo=FALSE, fig.width=5, fig.height=3.5} 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")

## Hypothesis tests

### Conditions

Since we dealing with subsets of n <= 30 our standard error estimate will generally not be accurate. Therefore, we will use the t di

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, we assume that the guinea pigs were randomly assigned to one of the groups (guarantees independence) and that they were sam

### Differences in supplement types

First, we will check if the observed difference in tooth length means between pigs who received their dose using orange juice and th

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

```{r echo=FALSE}
t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE)
tidy(t_diff_supp)
```

With a p-value of `r round(t_diff_supp$p.value, 3)` which is larger than the significance value, 0.05, we fail 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 `r round(t_diff_supp$conf.int[1:2], 3)` because it includes 0.

## Differences in dose levels

Since we 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.

```{r echo = FALSE} 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)

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

Interpreting the first row of the table for illustriation:

We are 95% confident that the average tooth length of guinea pigs who received a 2mg dose of vitamin C is on average `r round(pairwi

# Conclusions

We were able to show that the observed difference of tooth length mean across supplement types is statistically not different from 0

However, we were able to highlight the fact that tooth length indeed varies across groups of dose levels and that all of those diffe

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

## Appendix (Source Code)

This section includes the source code which was used for generating the results of the analysis. In addition you may check out the c

```{r eval=FALSE}
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)
```

```
```