# An Investigation of Tooth Growth in Guinea Pigs

*Project by Vanshaj S.*

## Introduction

In this "study," we're going to take a look at a sample of guinea pig tooth growth information and try to figure out whether certain doses or/of certain supplements are better at increasing growth than others.

## Required Libraries

For ease (and beauty) of graphing, let's use the `ggplot2` library with `gridExtra` (for the `grid.arrange()` function).

```
library(ggplot2)
library(gridExtra)

library(datasets)
```

## Setup

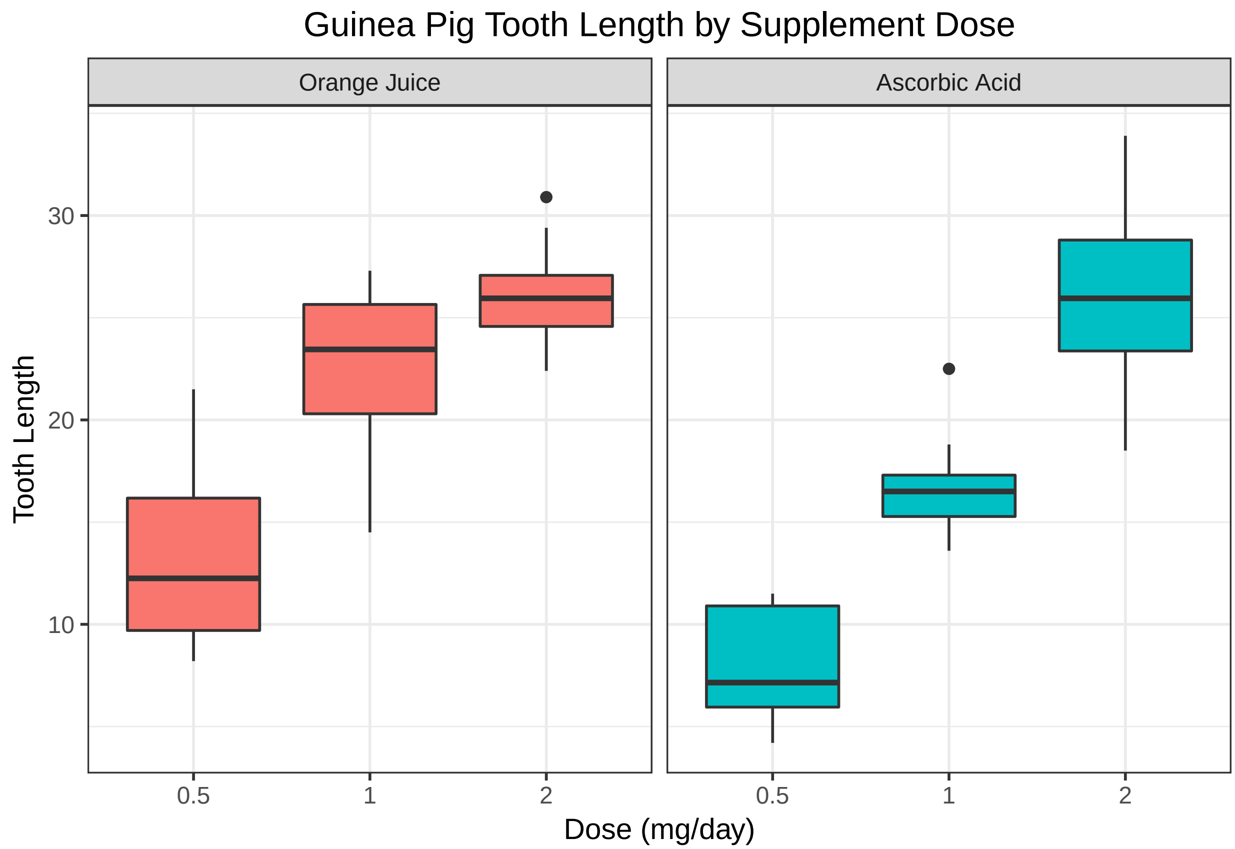
To ensure reproducibility, let's set a random seed. We'll also define our variables for future use.

```
teeth <- ToothGrowth      # a local copy won't hurt
set.seed(8243)            # reproducibility

steelblue <- "steelblue"  # a great fill color
orange <- "#c56000"       # a great line color
```

## EDA

Let's begin by doing some exploratory data anlysis.

```
levels(teeth$supp) <- c("Orange Juice", "Ascorbic Acid")

ggplot(teeth, aes(x = factor(dose), y = len)) +
  facet_grid(. ~ supp) +
  geom_boxplot(aes(fill = supp), show.legend = F) +
  theme_bw() +
  labs(title = "Guinea Pig Tooth Length by Supplement Dose",
       x = "Dose (mg/day)", y = "Tooth Length") +
  theme(plot.title = element_text(hjust = 0.5))
```



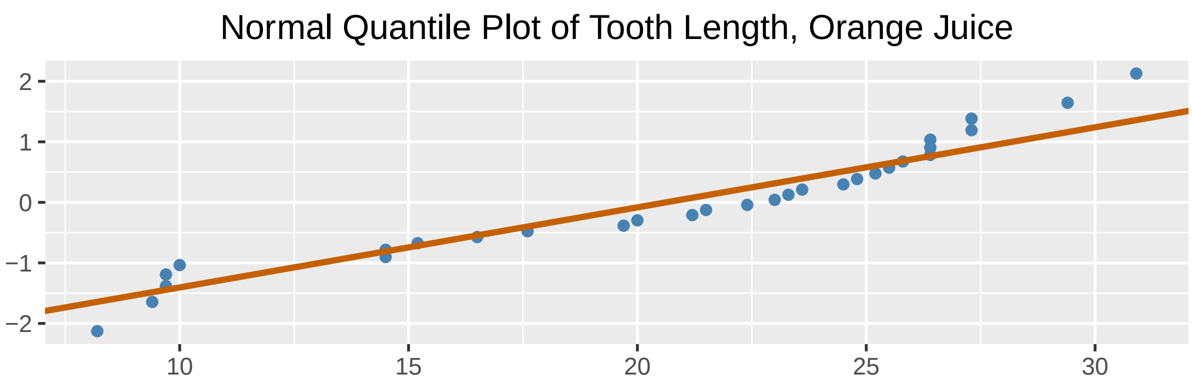

At first sight, it appears that orange juice is more effective in increasing tooth length across a 0.5 and 1.0 milligram/day dose. Ascorbic acid seems to have a greater range of effect when the dose is 2.0 milligrams/day. Regardless of the type of supplement, tooth growth seems to increase with higher doses.

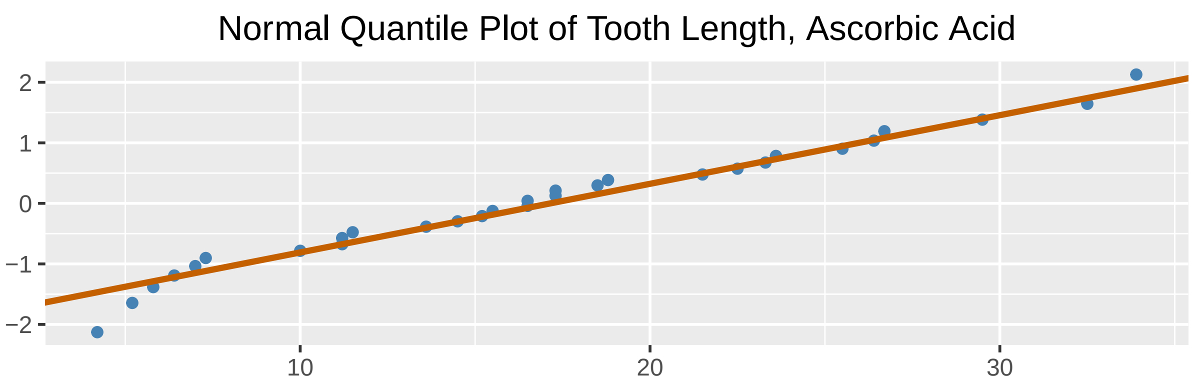
### Normality

Before conducting our hypothesis tests, let's perform an arbitrary check for normality on the datasets by supplement. For concision, we won't perform the same tests by dose.

```
x <- qnorm(c(0.25, 0.75))
y <- quantile(teeth[teeth$supp == "Orange Juice",]$len, c(0.25, 0.75))

slope <- diff(y)/diff(x)
int <- y[1L] - (slope * x[1L])

d <- data.frame(teeth[teeth$supp == "Orange Juice",]$len)

plot.oj <- ggplot(d, aes(sample = teeth[teeth$supp == "Orange Juice",]$len,
                         col = I(steelblue))) + geom_qq() +
  geom_abline(slope = slope, intercept = int, col = I(orange), size = 1.1) +
  labs(title = "Normal Quantile Plot of Tooth Length, Orange Juice", x = "", y = "") +
  theme(plot.title = element_text(hjust = 0.5)) + coord_flip()

## the next bit was moved to the next page for readability
```

```
x <- qnorm(c(0.25, 0.75))
y <- quantile(teeth[teeth$supp == "Ascorbic Acid",]$len, c(0.25, 0.75))

slope <- diff(y)/diff(x)
int <- y[1L] - (slope * x[1L])

d <- data.frame(teeth[teeth$supp == "Ascorbic Acid",]$len)

plot.vc <- ggplot(d, aes(sample = teeth[teeth$supp == "Ascorbic Acid",]$len,
                         col = I(steelblue))) + geom_qq() +
  geom_abline(slope = slope, intercept = int, col = I(orange), size = 1.1) +
  labs(title = "Normal Quantile Plot of Tooth Length, Ascorbic Acid", x = "", y = "") +
  theme(plot.title = element_text(hjust = 0.5)) + coord_flip()

grid.arrange(plot.oj, plot.vc, nrow = 2)
```




The data appears to match the approximate lines pretty well, so it's probably safe to assume that the datasets are both fairly normal. We'll also pretend that subsets of the data (i.e. by dose) are also normal. Let's begin.

### Test #1 – Is There a Difference Between the Growth Provided by the Supplements?

Our null hypothesis here is that there is *no* difference between the growth provided by orange juice and ascorbic acid, regardless of the dose. The alternate, it follows, is that there is *a* difference. We'll use a two-sided t-test.

```
by.supp <- t.test(len ~ supp, data = teeth)
by.supp
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                    20.66333                    16.96333
```

The 95% confidence interval ranges from -0.1710156 to 7.5710156, a range which includes 0. Furthermore, the p-value is 0.0606345, greater than a 0.05 alpha level. For these two reasons, we fail to reject the null hypothesis – there appears to be a possibility that there is no difference between the growth provided by each type of supplement.

Not to be dejected by this failure, we shall now take an XKCD-style approach. Let's take a look at the different doses.

## Test #2 – Is There a Difference Between the Growth Provided by the Supplements for a 0.5 mg/day Dose?

Our null hypothesis here is that there is *no* difference between the growth provided by orange juice and ascorbic acid for a 0.5 mg/day dose. The alternate, it follows, is that there is *a* difference. We'll use a two-sided t-test.

```
by.dose.half <- t.test(len ~ supp, data = teeth[teeth$dose == "0.5",])
by.dose.half
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                       13.23                        7.98
```

The 95% confidence interval ranges from 1.7190573 to 8.7809427, a range which does not include 0. Furthermore, the p-value is 0.0063586, less than a 0.05 alpha level. For these two reasons, we successfully reject the null hypothesis – there appears to be a possibility that there is indeed a difference between the growth provided by each type of supplement for a 0.5 mg/day dose.

## Test #3 – Is There a Difference Between the Growth Provided by the Supplements for a 1.0 mg/day Dose?

Our null hypothesis here is that there is *no* difference between the growth provided by orange juice and ascorbic acid for a 1.0 mg/day dose. The alternate, it follows, is that there is *a* difference. We'll use a two-sided t-test.

```r
by.dose.one <- t.test(len ~ supp, data = teeth[teeth$dose == "1",])
by.dose.one
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                       22.70                       16.77
```

The 95% confidence interval ranges from 2.8021482 to 9.0578518, a range which does not includes 0. Furthermore, the p-value is 0.0010384, less than a 0.05 alpha level. For these two reasons, we successfully reject the null hypothesis – there appears to be a possibility that there is indeed a difference between the growth provided by each type of supplement for a 1.0 mg/day dose.

### Test #4 – Is There a Difference Between the Growth Provided by the Supplements for a 2.0 mg/day Dose?

Our null hypothesis here is that there is *no* difference between the growth provided by orange juice and ascorbic acid for a 2.0 mg/day dose. The alternate, it follows, is that there is *a* difference. We'll use a two-sided t-test.

```r
by.dose.two <- t.test(len ~ supp, data = teeth[teeth$dose == "2",])
by.dose.two
```

```
##
##  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 Orange Juice mean in group Ascorbic Acid
##                       26.06                       26.14
```

The 95% confidence interval ranges from -3.7980705 to 3.6380705, a range which includes 0. Furthermore, the p-value is 0.9638516, greater than a 0.05 alpha level. For these two reasons, we fail to reject the null hypothesis – there appears to be a possibility that there is no difference between the growth provided by each type of supplement for a 2.0 mg/day dose.

## Conclusions

Because orange juice was evaluated first (so the calculation was in the order "OJ" - "VC"), we can conclude from our last three tests that orange juice does a better job at tooth growth at a 0.5 and 1.0 mg/day dose. However, for a 2.0 mg/day dose, it is not distinguishable which one works better. Furthermore, if we choose to ignore the dose level, it is again not distinguishable which one works better.

Once again, we assumed that the datasets are normal (even when split up by dose) and, more importantly, that there are no confounding variables, i.e. external variables that may affect tooth growth (which, of course, is probably not true).