

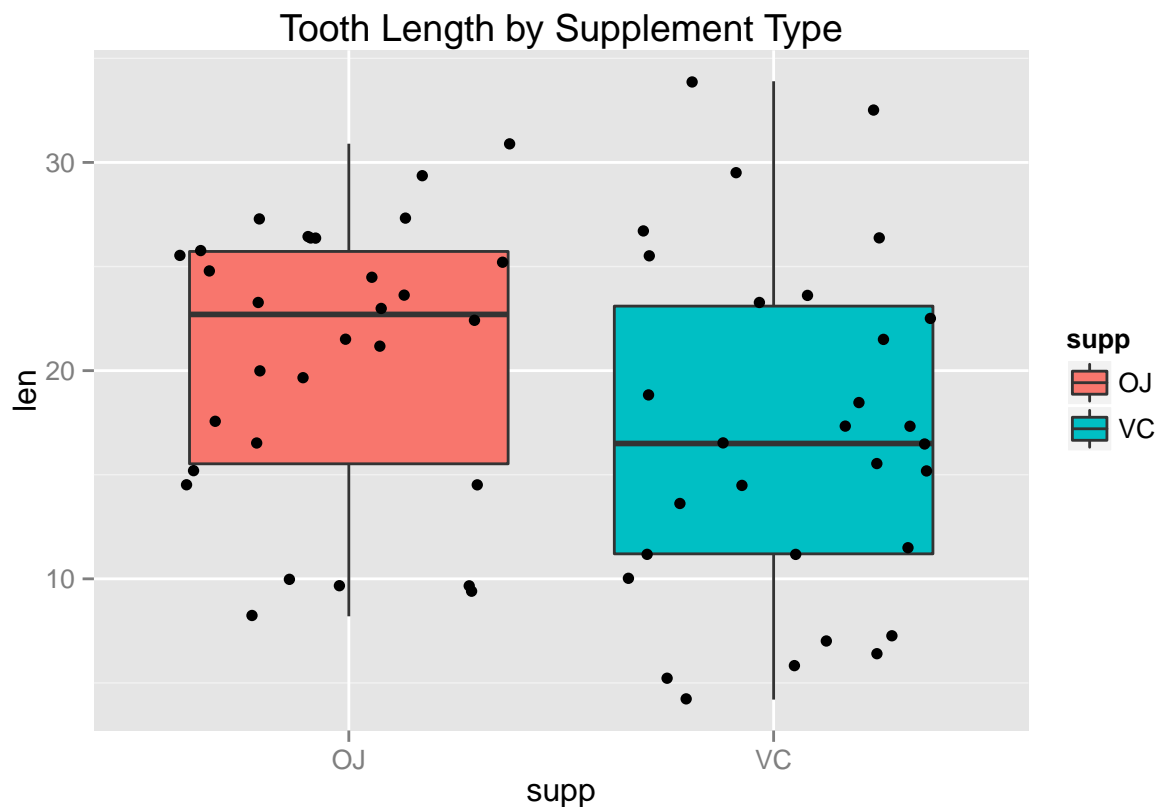
Tooth Growth Analysis

Ran Ju

Exploratory Data Analysis

```
# Load library and data
library(ggplot2)
data(ToothGrowth)
tg <- ToothGrowth
# Fix the seed for reproducibility
set.seed(344344)
```

```
# Box plot of len against supp
ggplot(data = tg, aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp)) +
  geom_jitter() +
  ggtitle("Tooth Length by Supplement Type")
```

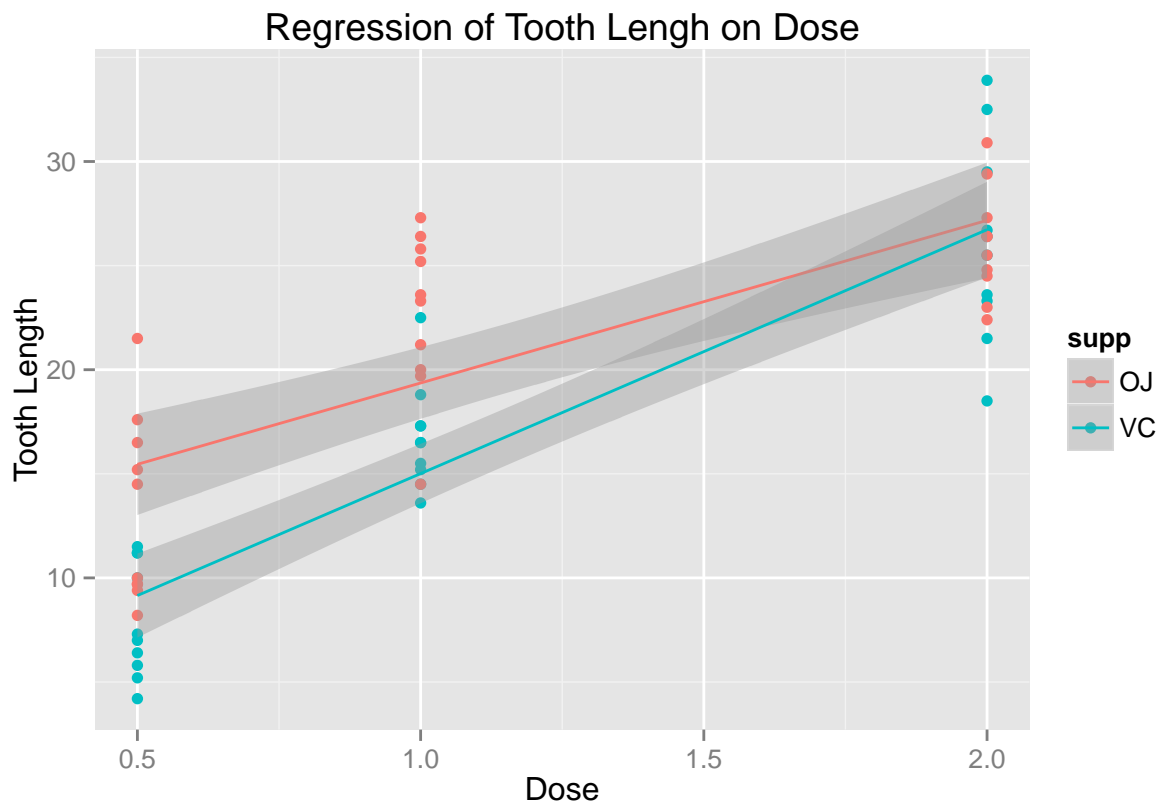


```
# Box plot of len against dose
ggplot(data = tg, aes(x = factor(dose), y = len)) +
  geom_boxplot(aes(fill = factor(dose))) +
```

```
geom_jitter() +
ggtitle("Tooth Length by Dose")
```



```
qplot(dose, len, data=tg, geom=c("point", "smooth"),
      method="lm", formula=y~x, color=supp,
      main="Regression of Tooth Length on Dose",
      xlab="Dose", ylab="Tooth Length")
```

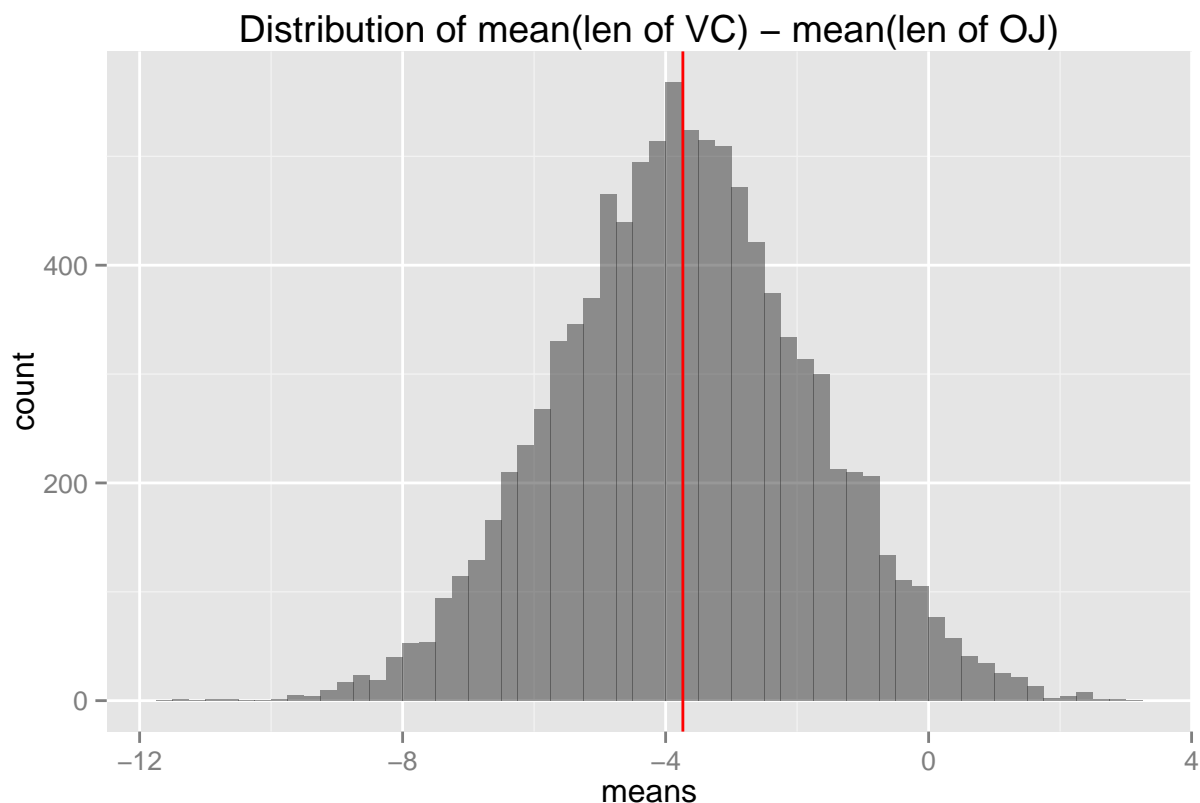


It seems that orange juice is better for tooth length than ascorbic acid, and the more vitamin C the higher the tooth length.

Is tooth length related to supplement?

Define H_0 as $E[\text{len}|\text{supp} = \text{OJ}] = E[\text{len}|\text{supp} = \text{VC}] = E[\text{len}]$ with $\alpha = 0.05$. Use resampling to create 10000 simulations, and compute the difference of mean of OJ and VC for each sample.

```
n = dim(tg)[1]
B = 10000
means <- NULL
for (i in 1:B) {
  # Sample with replacement
  ri <- sample.int(n, n, replace = T)
  t <- tg[ri, ]
  # Compute the difference of the mean values
  means <- c(means, mean(t[t[, "supp"] == "VC", "len"]) - mean(t[t[, "supp"] == "OJ", "len"]))
}
ggplot() +
  geom_histogram(aes(x = means), alpha = 0.5, binwidth = 0.25) +
  geom_vline(aes(xintercept = median(means)), color = "red") +
  ggtitle("Distribution of mean(len of VC) - mean(len of OJ)")
```



```
# Compute the median
median(means)
```

```
## [1] -3.737445
```

The plot shows the distribution of the means (the vertical line is the median), and it is less than 0 with probability

```
# Compute the weight of mean >= 0
sum(means >= 0) / B
```

```
## [1] 0.0285
```

Because the probability is less than $\alpha = 0.05$, we reject H_0 .

Is tooth growth related to dose?

We perform a t-test between each pair of dose levels.

```
t05 <- tg[tg$dose == 0.5, "len"]
t1 <- tg[tg$dose == 1, "len"]
t2 <- tg[tg$dose == 2, "len"]
```

```

# Compare dose = 0.5 and 1
t.test(t05, t1, alternative = "less")

##
## Welch Two Sample t-test
##
## data: t05 and t1
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -6.753323
## sample estimates:
## mean of x mean of y
##      10.605      19.735

# Compare dose = 1 and 2
t.test(t1, t2, alternative = "less")

##
## Welch Two Sample t-test
##
## data: t1 and t2
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##      19.735      26.100

# Compare dose = 0.5 and 2
t.test(t05, t2, alternative = "less")

##
## Welch Two Sample t-test
##
## data: t05 and t2
## t = -11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -13.27926
## sample estimates:
## mean of x mean of y
##      10.605      26.100

```

The p-values are all well below $\alpha = 0.05$, hence we conclude that VC dose has a positive effect on tooth growth, assuming *len* has a normal distribution in each group.

Conclusion

Tooth length is higher among the individuals fed with orange juice than ascorbic acid, and VC dose has a positive correlation with tooth growth. However, more analysis is needed to quantify the relationship, which requires regression analysis.