investigating the exponential distribution

Daniel Aklilu

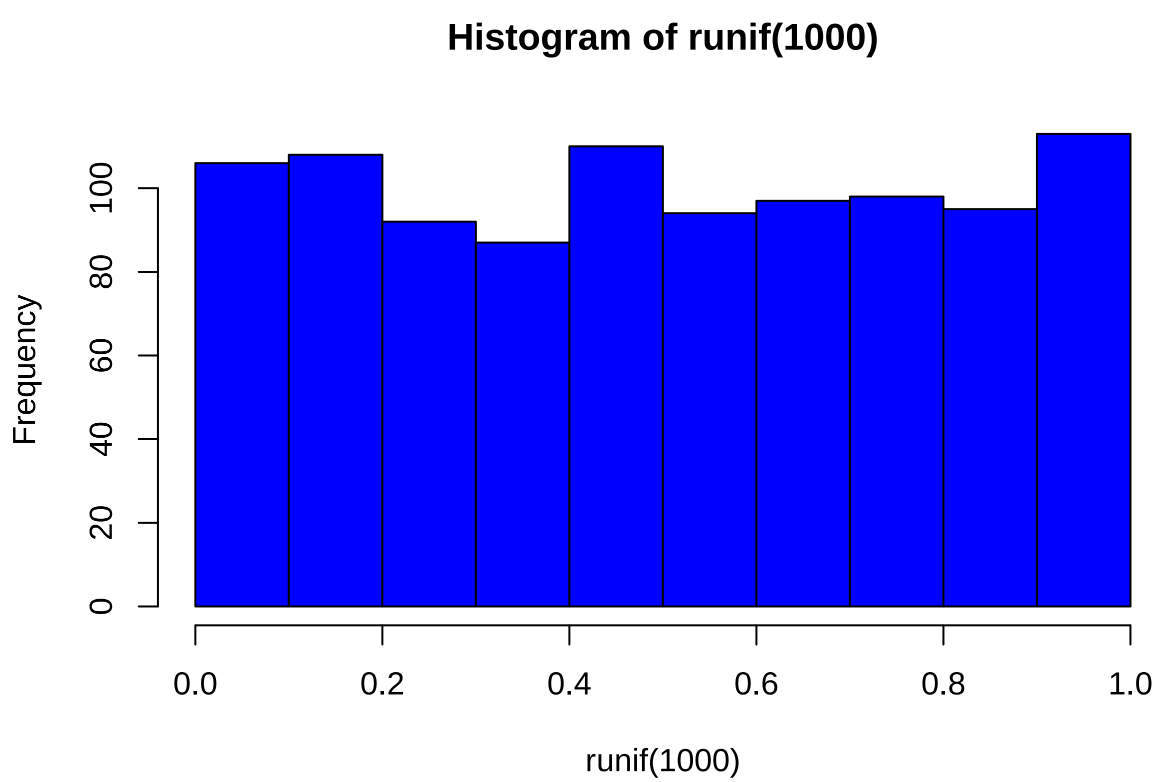
2025-07-23

Overview:

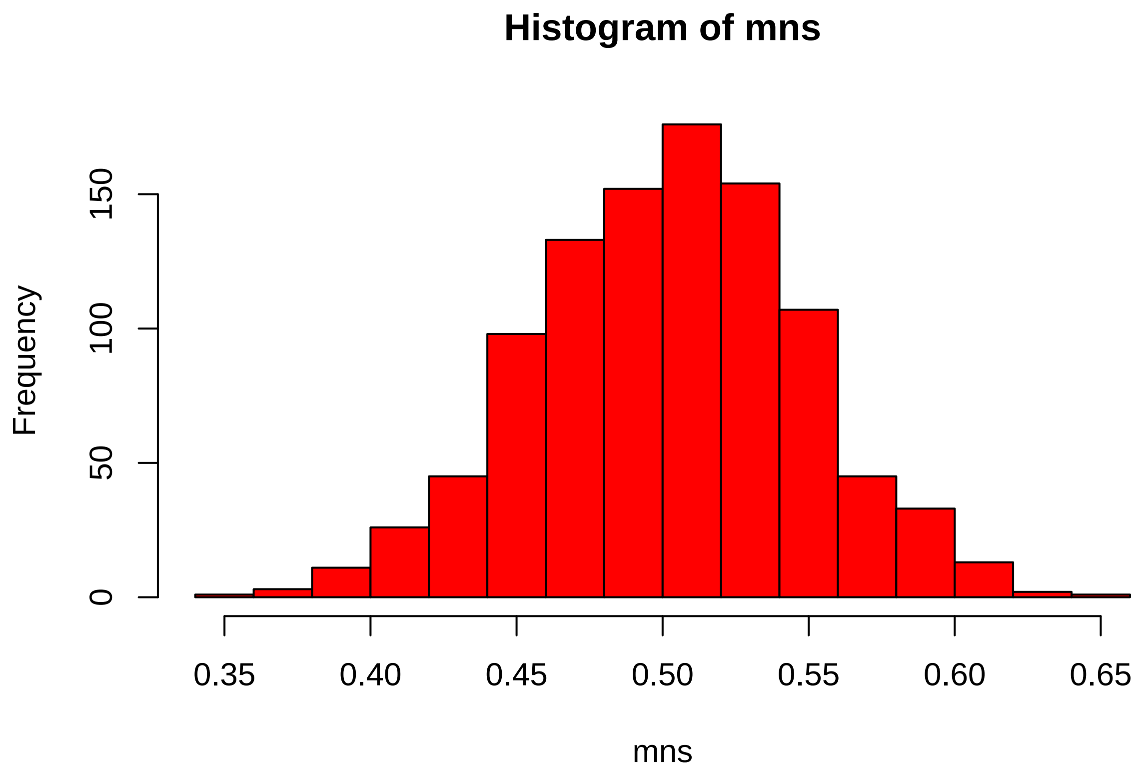
In this project we will be exploring statistical inference through two components: a simulation study and an analysis of real-world data. In the first part, we use simulated exponential data to demonstrate the Central Limit Theorem. In the second part, we analyzw the ToothGrowth dataset using exploratory plots and hypothesis test to assess the effects of supplement type and dosage on tooth growth.

Histogram of 1000 random samples, and the averages of 40 exponentials

```
#get histogram of 1000 random samples
hist(runif(1000), col = "blue")
```



```r
#set mns equal to empty vector
mns = NULL
#itrates though 1000 samples,
#and generates 40 random valjues in each iteration,
#Computes their mean, and appends it to the vector mns.
for(i in 1:1000)mns = c(mns, mean(runif(40)))
# Plot a histogram of 1000 sample means
hist(mns, col = "red")
```



Simulate one Sample of 40 exponentials

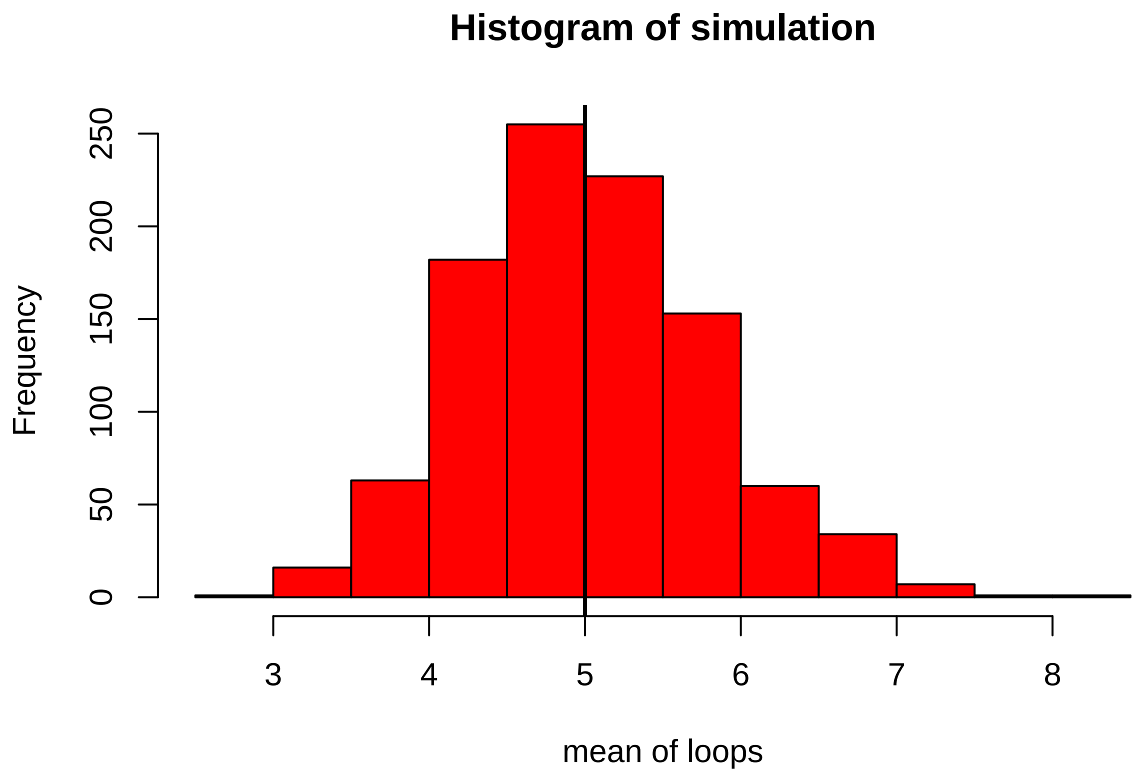
```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

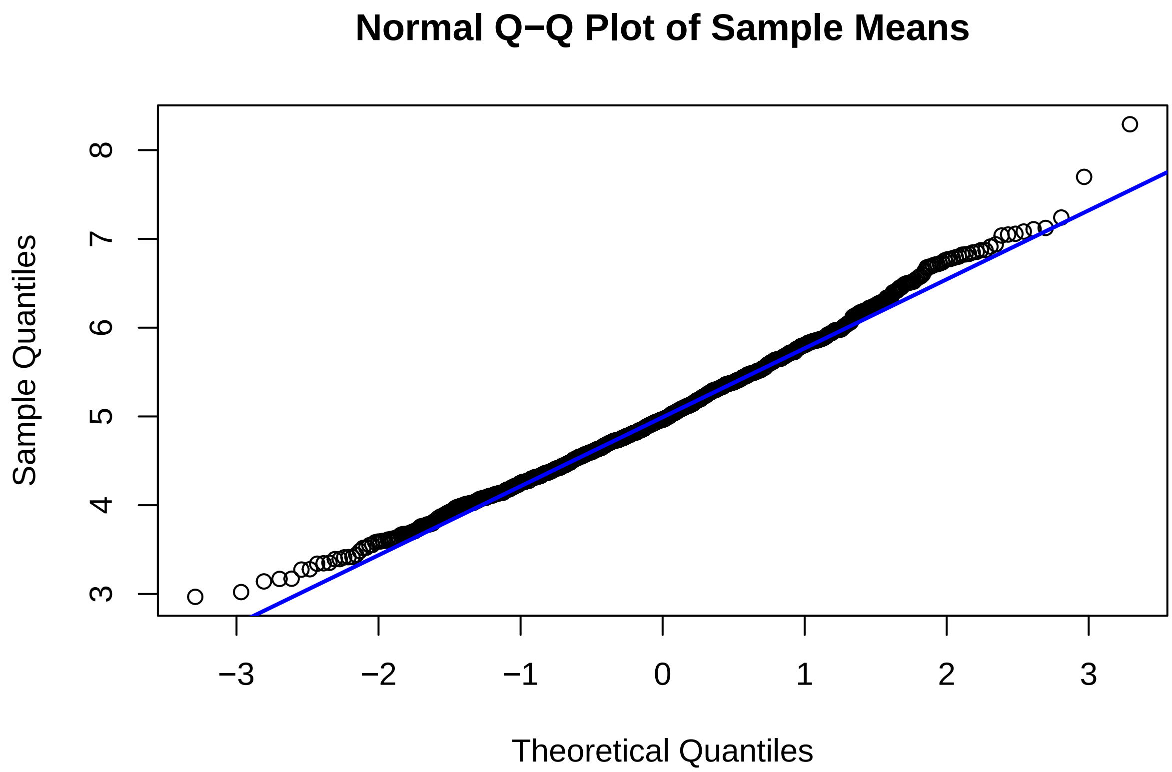
```r
library(dplyr)
library(ggplot2)
#install.packages("tinytext", repos = "https://cloud.r-projects.org/")
#tinytex::install_tinytex()
tinytex::tinytex_root()
# varaible declaration
lambda <- 0.2
simulation <- 1000
size <- 40

# Simulate and Store Sample Means
mean_size <- rexp(size, rate = lambda)
# Simulates 1000 times and stores the mean of each sample.
mean_loop <- sapply(1:simulation, function(i) mean(rexp(size,rate = lambda)))
```

Simulation of Sample Means and Normality Check

```r
#Check the sample mean
sim_avg <- mean(mean_loop)

#Check the sample variance
sam_var <- var(mean_loop)

#Plot the distribtion
hist(mean_loop, col = "red",
     main = "Histogram of simulation",
     xlab = "mean of loops")
# Add a vertical reference line at the
# true mean (e.g., population mean = 5)
abline(v = 5, col = "black", lwd =2)
```




```r
# Create a Q-Q plot to check for normality of the sample
# means. If points fall along the line, the
# distribution is approximately normal.
qqnorm(mean_loop, main= "Normal Q-Q Plot of Sample Means")
qqline(mean_loop, col = "blue", lwd =2)
```



Interpreting Q-Q plot

```
# returns the mean of sim_avg
  sim_avg
```

```
## [1] 5.021622
```

```
# returns the variance of sam_var
  sam_var
```

```
## [1] 0.6137434
```

Summary of simulation

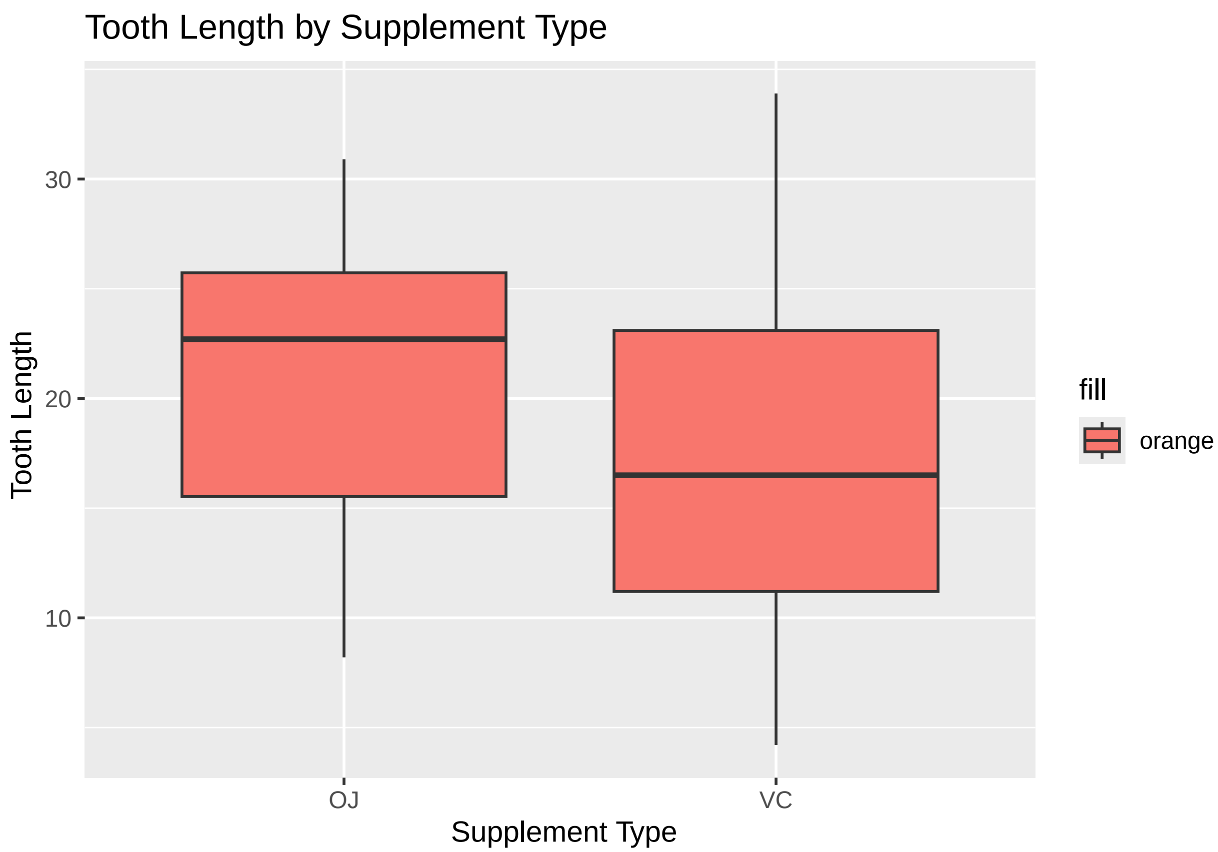
In this simulation, we generated 1000 sample means using the random samples of size 40 from an exponential distribution with a rate parameter of 0.2. The theortical mean of this distribution is 5, and the theoretical mean of this distribution is 5.036, and the theortical variance of the sample mean is 0.625. The simulated sample mean was approximately 5.037, and the variance was approximately 0.652 - both closely aligned with the expected theoretical values. The histogram of the sample means shows a symmetric, bell-shaped distribution centered near 5. Additionally, the Q-Q plot shows that the distribution of the sample means is about normal, which confirms the Central Limit Theorem.

Analyze tooth growth dataset

```r
head(ToothGrowth,10)
summary(ToothGrowth)
str(ToothGrowth)

var(ToothGrowth)

sapply(ToothGrowth[, c("len","dose")], var)
```

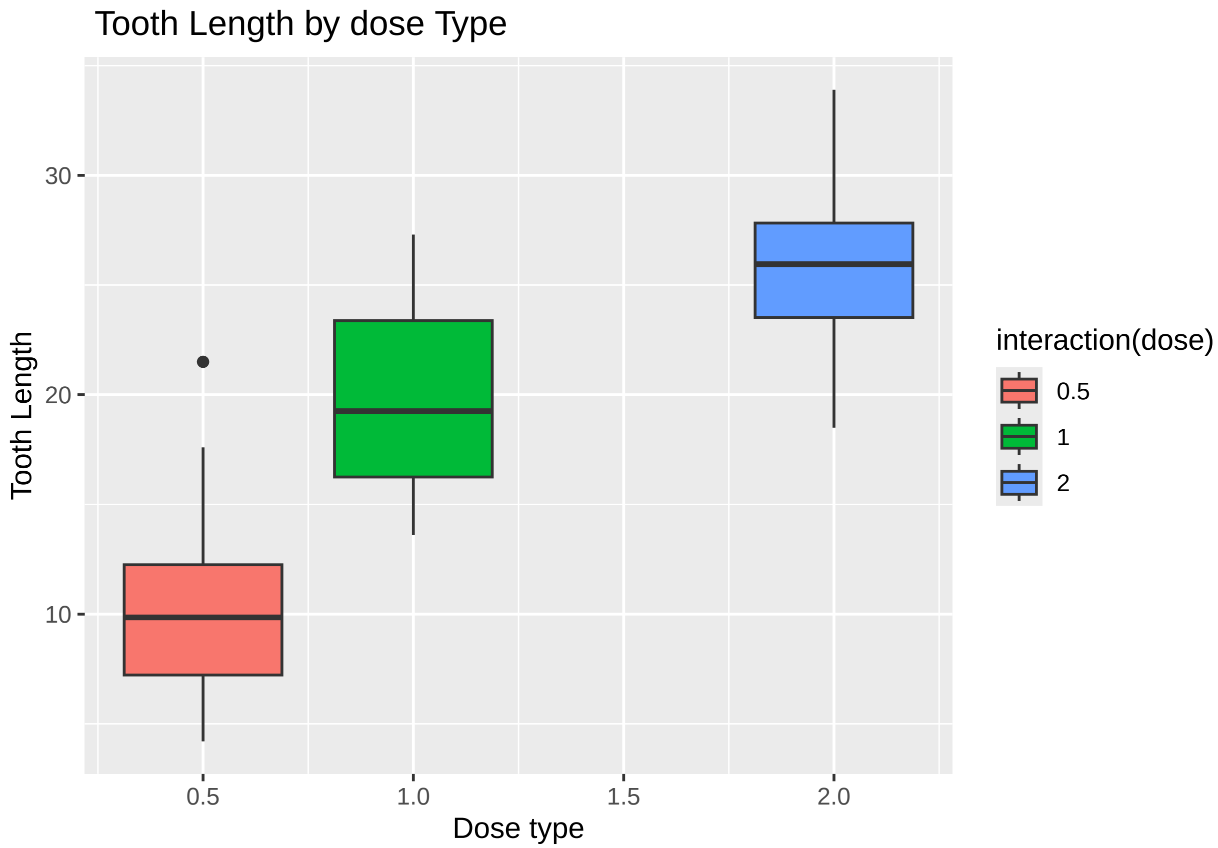
Plotting Tooth Length by Supplement Type

```r
# This plot compares overall tooth length between the two supplement groups (OJ and VC)
# ignoring the effect of dosage.
ggplot(ToothGrowth, aes(x= supp, y = len, fill = "orange")) + geom_boxplot() + labs(title = "Tooth Leng
```



Plotting Tooth Length by Dose Type

```r
# This plot shows how tooth length varies by both supplement type(OJ vs VC)
# and dosage at (0.5, 1.0, 2.0 mg/day), with color representing supplement type.
ggplot(ToothGrowth, aes(x = dose, y = len, fill = interaction(dose))) + geom_boxplot() + labs(title = "
```



Results of Boxplot: Tooth Length by Supplement Type

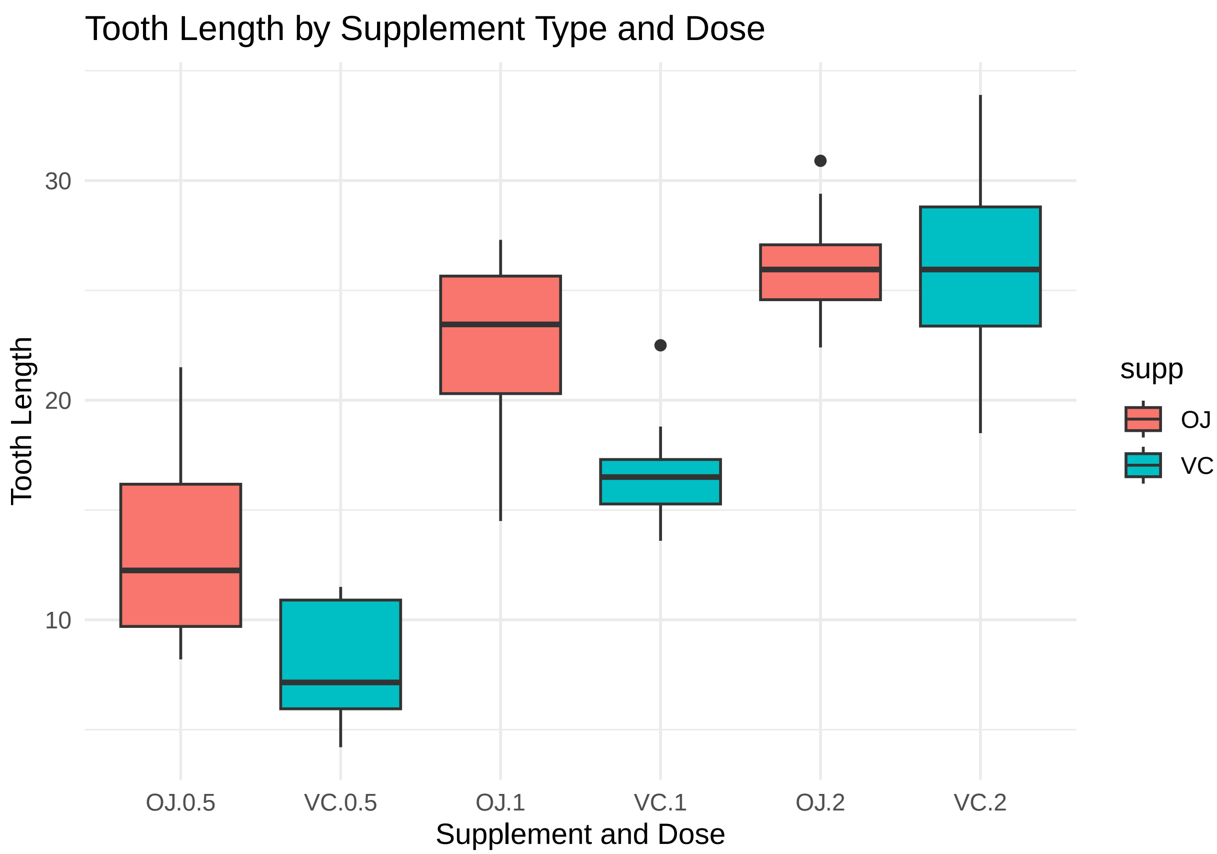
The boxplot compares tooth length between the two supplement types: OJ(Orange Juice) and VC (Vitamin C). Overall, OJ appears to result in slightly longer tooth growth compared to VC, as indicated by a higher median and upper quartile. However, there is noticeable overlap in variability between the two groups, suggesting the difference in means is not drastic without considering dose. This plot alone suggest OJ may be more effective than VC, but additional analysis is needed across different dose levels to understand this trend.

Results of Boxplot: Tooth Length by Dose Type

The boxplot illustrates that tooth length increases with higher doses of vitamin C. The median tooth length is lowest at 0.5 mg/day and highest at 2.0 mg/day, indicating a clear positive relationship between dosage and tooth growth. Additionally, the variability in tooth length appears to decrease slightly at the highest dose, suggesting more consistent outcomes at 2.0 mg/day. This trend supports the hypothesis that increased vitamin C dosage enhances tooth development.

Plotting Tooth Length by Supplement Type and Dose

```
# This plot shows how tooth length varies by both supplement type (OJ vs VC)
# and dosewa at (0.5, 1.0, 2.0 mg/day), with color representing supplement type.
ggplot(ToothGrowth, aes(x = interaction(supp, dose), y = len, fill = supp)) + geom_boxplot() + labs(titl
```



Results of Boxplot: Tooth length by Supplement and Dose

The Boxplot shows that tooth length generally increases with higher doses of vitamin C, regardless of supplement type. There is noticeable separation between the groups, with the 2 mg/ day dose resulting in the highest tooth growth. The variability also appears to decrease at higher doses.

Filtered Data by dose and ran two sample t-test

```r
#I filtered the data by dose 0.5 mg/day
filtered_Data <- filter(ToothGrowth, dose == 0.5)
t.test(len ~ supp, filtered_Data)
```

```
##
##  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 between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
```

```
##            13.23             7.98
```

```r
# I filtered the data by dose at 0.1 mg/day
filtered_Data2 <- filter(ToothGrowth, dose == 1)
t.test(len ~ supp, filtered_Data2 )
```

```
##
##  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 between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
# I filtered the data by dose at 0.2 mg/dau
filtered_Data3 <- filter(ToothGrowth, dose == 2)
t.test(len~supp,filtered_Data3)
```

```
##
##  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 between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

```r
# I am filtering by OJ
OJ_filtered <- filter(ToothGrowth, supp == "OJ")
# Compare len between doses 0.5 and 2
t.test(len ~ dose,data =
         filter(OJ_filtered, dose %in%  c(0.5, 2)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -7.817, df = 14.668, p-value = 1.324e-06
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -16.335241  -9.324759
## sample estimates:
## mean in group 0.5   mean in group 2
##             13.23             26.06
```

Summary of Findings

- At dose = 0.5 mg/day, supplement OJ resulted in significantly greater tooth growth than VC (p-value = 0.006).
- At dose = 1.0 mg/day, supplement OJ again outperformed VC(p- value = 0.001).
- At dose = 2.0 mg/day, there was no significant difference in tooth length between OJ and VC (p-value = 0.964).
- When comparing dose 0.5 vs 2.0 within the OJ group, the higher dose significantly increased tooth length (p-value < 0.001).

  The results suggest that both the type of supplement and the dose level influence tooth growth, but the difference between supplement types disappears at higher doses.

Conclusion

This analysis of the ToothGrowth dataset demonstrates that both supplement type and dose level significantly influence tooth length in guinea pigs. At lower doses (0.5 and 1.0 mg/day), OJ consistently led to greater tooth growth than VC, with statistically significant differences. However, at the highest dose 2.0 mg/day, the difference between the two supplements was negligible, suggesting that dose becomes the dominant factor at higher levels. These results support the hypothesis that vitamin C dosage has a effect on tooth growth than supplement delivery methods at higher concentrations. We reject the null hypothesis for comparisons(0.5 and 1.0 doses), showing a significant difference in tooth growth, except for the highest dose (2.0 mg/day) between OJ and VC, where you fail to reject the null hypothesis.