Final Project - ASDS 5301

Study on the growth of the odontoblast of the incisor teeth as a criterion of Vitamin C intake of the Guinea Pigs using Two Way ANOVA

Submitted to:
Dr. Mahmoud Jawad
Applied Statistics and Data Science
College of Science
University of Texas at Arlington

Report Prepared By:
Abhishek Goudar (1001916610)
University of Texas at Arlington

Jeyasooriya Saravanan (1002186838)
University of Texas at Arlington

INTRODUCTION

During the Second World War, there was a notable concern regarding the provision of vitamin C to soldiers. In this broader context, researchers explored the effects of both ascorbic acid and orange juice in animal studies. A specific investigation within this domain was the Crampton's study, wherein 60 guinea pigs were administered a vitamin C dietary supplement at one of three doses (0.5, 1, or 2 mg/day). The supplement was delivered in one of two ways—either as ascorbic acid or through orange juice.

This experiment involved sixty 28-day-old guinea pigs that were administered a dietary supplement of vitamin C at one of three doses (0.5, 1, or 2 mg/day) delivered in either ascorbic acid or orange juice. Each combination of dose and delivery was represented by ten guinea pigs, evenly split between males and females. After being on this diet for 42 days, the guinea pigs were euthanized, and their incisors were extracted and sectioned to measure the length of odontoblasts. To obtain a single value for each animal, the lengths, which could include multiple measurements per animal, were averaged. The primary focus of the study is to understand how the dose and method of delivery influence the average length of incisor odontoblasts, measured in microns.

DATA OVERVIEW

```
df_tooth_growth <- data.frame(ToothGrowth)

# Structure of the dataset
str(df_tooth_growth)
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
head(df_tooth_growth[c(28, 29, 30, 31, 32, 33), ],
     n = 6)
##     len supp dose
## 28 21.5   VC  2.0
## 29 23.3   VC  2.0
## 30 29.5   VC  2.0
## 31 15.2   OJ  0.5
## 32 21.5   OJ  0.5
## 33 17.6   OJ  0.5
```

DESCRIPTIVE ANALYSIS

```
##Dataset summary

summary(df_tooth_growth)
```

```
##       len          supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000
##Descriptive stats in conjunction to both factors
require(dplyr)
library(dplyr)
group_by(df_tooth_growth, supp, dose) %>%
  summarise(
    count = n(),
    mean = mean(len, na.rm = TRUE),
    sd = sd(len, na.rm = TRUE),
    .groups = "drop"
  )
## # A tibble: 6 × 5
##   supp   dose count  mean    sd
##   <fct> <dbl> <int> <dbl> <dbl>
## 1 OJ      0.5    10 13.2   4.46
## 2 OJ      1      10 22.7   3.91
## 3 OJ      2      10 26.1   2.66
## 4 VC      0.5    10  7.98  2.75
## 5 VC      1      10 16.8   2.52
## 6 VC      2      10 26.1   4.80
```

- The variable 'len' (length) has a wide range from a minimum of 4.20 to a maximum of 33.90, suggesting a diverse distribution of lengths in the dataset.
- The 'supp' variable indicates that the dataset consists of two types of supplements, "OJ" and "VC," each with 30 observations. This suggests a balanced distribution between the two supplement types.
- The 'dose' variable ranges from 0.500 to 2.000, with the majority of doses concentrated below 2.000. The median dose is 1.000, indicating that half of the observations have a dose at or below this level.

BOXPLOTS

```
## Boxplot for Delivery Methods of Supplement
par(mar = c(5, 5, 2, 5))
boxplot(len~supp, data=df_tooth_growth,
        notch=FALSE,
  col=(c("lightcoral","lightgreen")),
  main="Delivery Methods for Vitamin C Supplements",
  xlab="Supplement Type (Asorbic Acid/Orange Juice)",
```

```
  ylab="Length of Odontoblast (Microns)",
  width = c(0.5, 1.5)*2,
  height = c(0.5, 1.5)
)
```

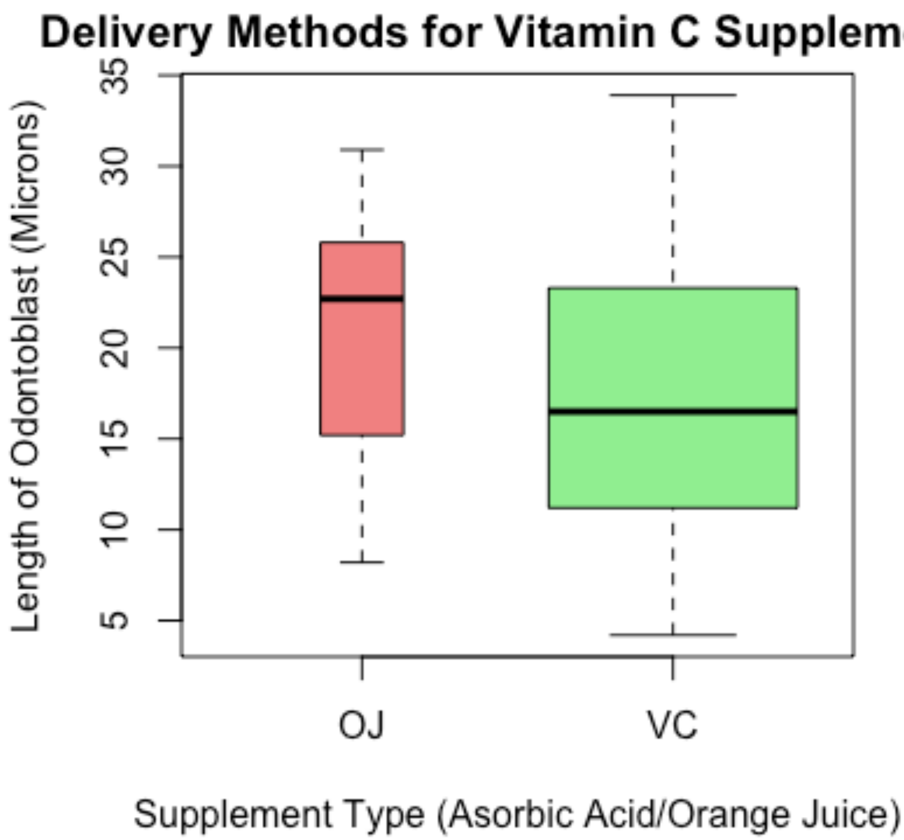


Figure 1. Delivery Methods of Vitamin C Supplements

Figure 1. shows the boxplots of Length of Odontoblast for each supplement independent of dosage.

Notice the median length, IQR (representing middle 50% of data) and size of boxplot (indicating spread or variability in the groups).

```
boxplot(len~dose, data=df_tooth_growth, notch=FALSE,
  col=(c("lightcoral","lightgreen")),
  main="Three Dose Levels of Vitamin C (0.5, 1, and 2)",
  xlab="Numeric Dose in milligrams/day",
  ylab="Length of Odontoblast (Microns)"
  )
```

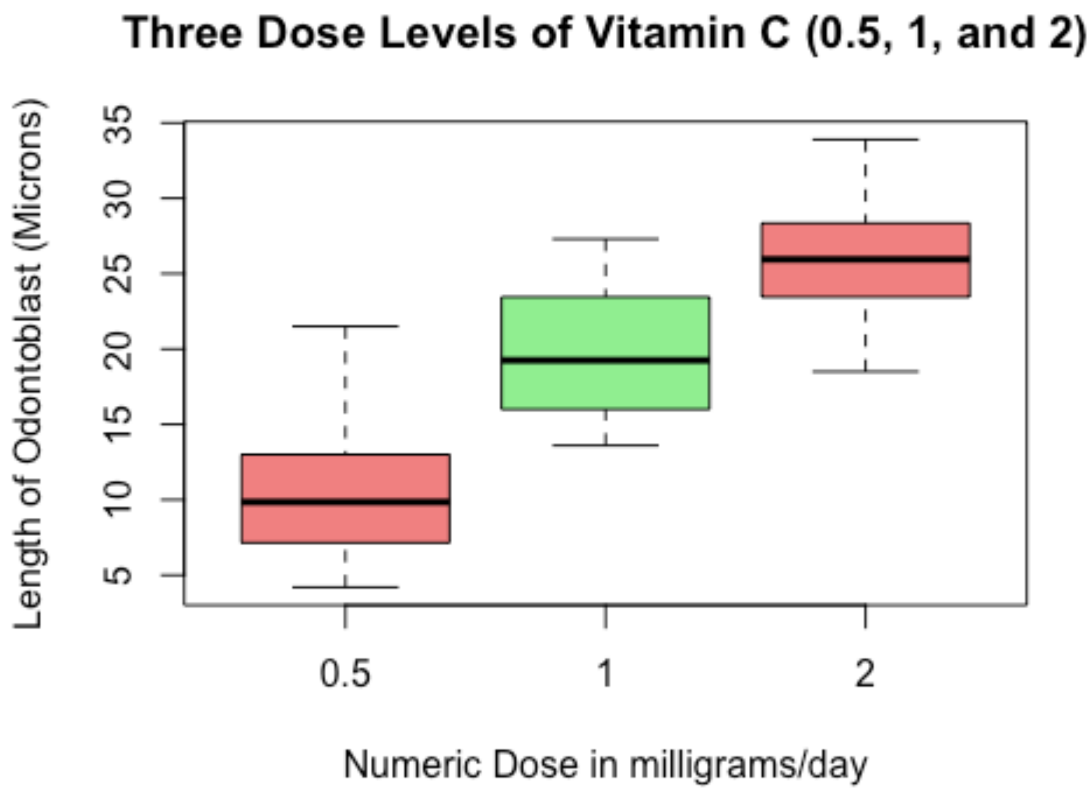


Figure 2. Boxplot of length of Odontoblast for each dosage.

Above figure shows that range of length of Odontoblast increases with dosage.

```
boxplot(len~supp*dose, data=df_tooth_growth, notch=FALSE,
  col=(c("lightcoral","lightgreen")),
  main="Odontoblast Growth",
  xlab="Suppliment and Dose",
  ylab="Length of Odontoblast (Microns)")
```

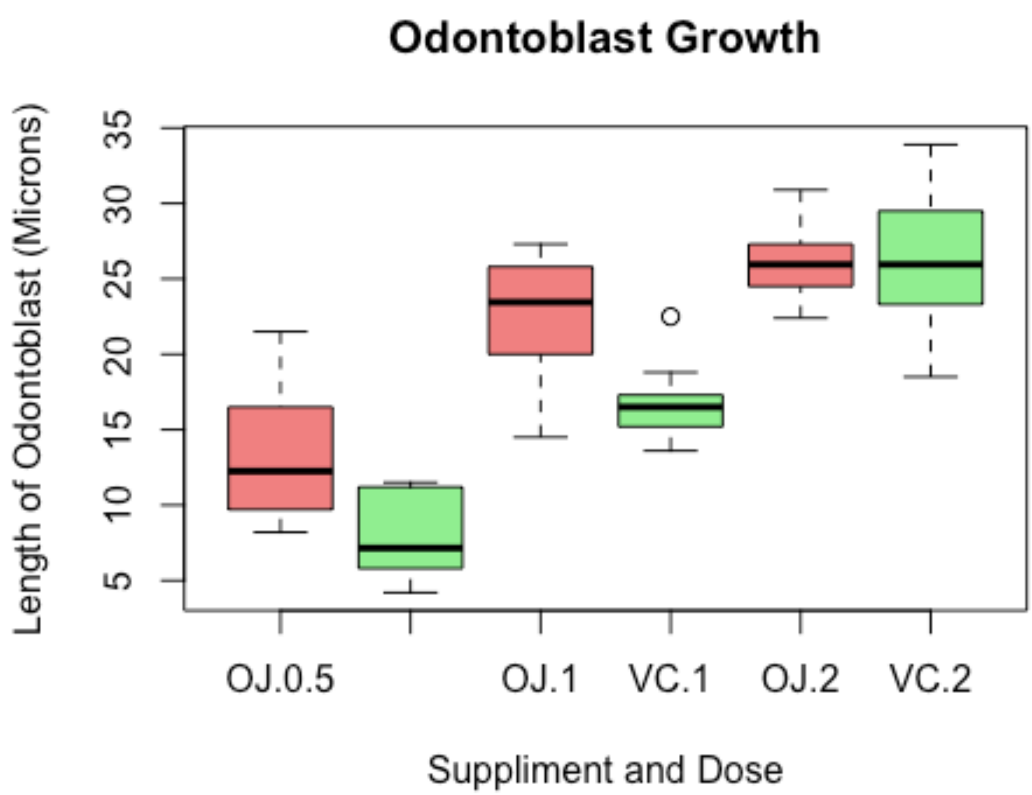


Figure 3. Boxplot of length of Odontoblast for each supplement with dosage.

This Boxplot indicates that when the dosage is set at 0.5 or 1.0 mg/day, the delivery of vitamin C through orange juice appears to be a more effective method, given that longer odontoblasts are considered advantageous. However, the significance of the delivery method seems to diminish when the dose is increased to 2.0 mg/day.

MAIN EFFECTS PLOT

```
# Main Effects Plot for 'dose'
ggplot(df_tooth_growth, aes(x = dose, y = len, group = supp, color = supp)) +
  geom_line() +
  geom_point() +
  labs(title = "Main Effects Plot for Dose",
       x = "Dose (milligrams/day)", y = "Length of Odontoblast (Microns)") +
  scale_color_manual(values = c("lightcoral", "lightgreen")
                     ) +  # Set custom colors
  theme_minimal()
```

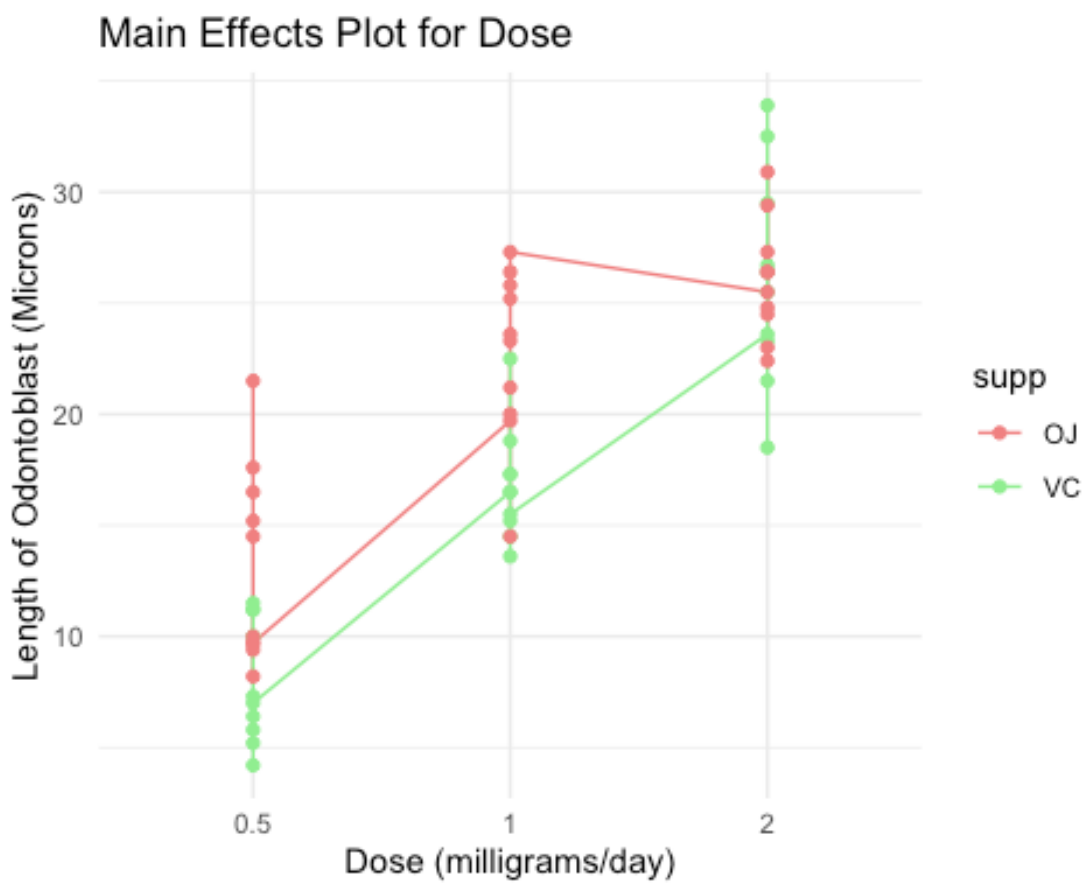


Figure 5. Main Effects Plot for Dose.

The main effects plot of length of odontoblast vs dose shows that the mean length of odontoblast increases with dose, regardless of the supplement type. The slope of the line is positive for both orange juice (OJ) and ascorbic acid (VC) supplements, indicating that a higher dose of vitamin C leads to longer odontoblasts. The average length of odontoblasts is greater for guinea pigs that received 1 mg/day or 2 mg/day of vitamin C compared to those that received 0.5 mg/day

```
means_data <- df_tooth_growth %>%
  group_by(supp, dose) %>%
  summarize(mean_len = mean(len),
            .groups="drop")

# Line plot of means
ggplot(means_data, aes(x = dose, y = mean_len, group = supp, color = supp)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "Mean Odontoblast Length for Each Supplement vs Dose",
       x = "Dose (milligrams/day)", y = "Mean Odontoblast Length (Microns)")
+
  scale_color_manual(values = c("lightcoral", "lightgreen")
```

```
                    ) +  # Set custom colors
  theme_minimal()
```

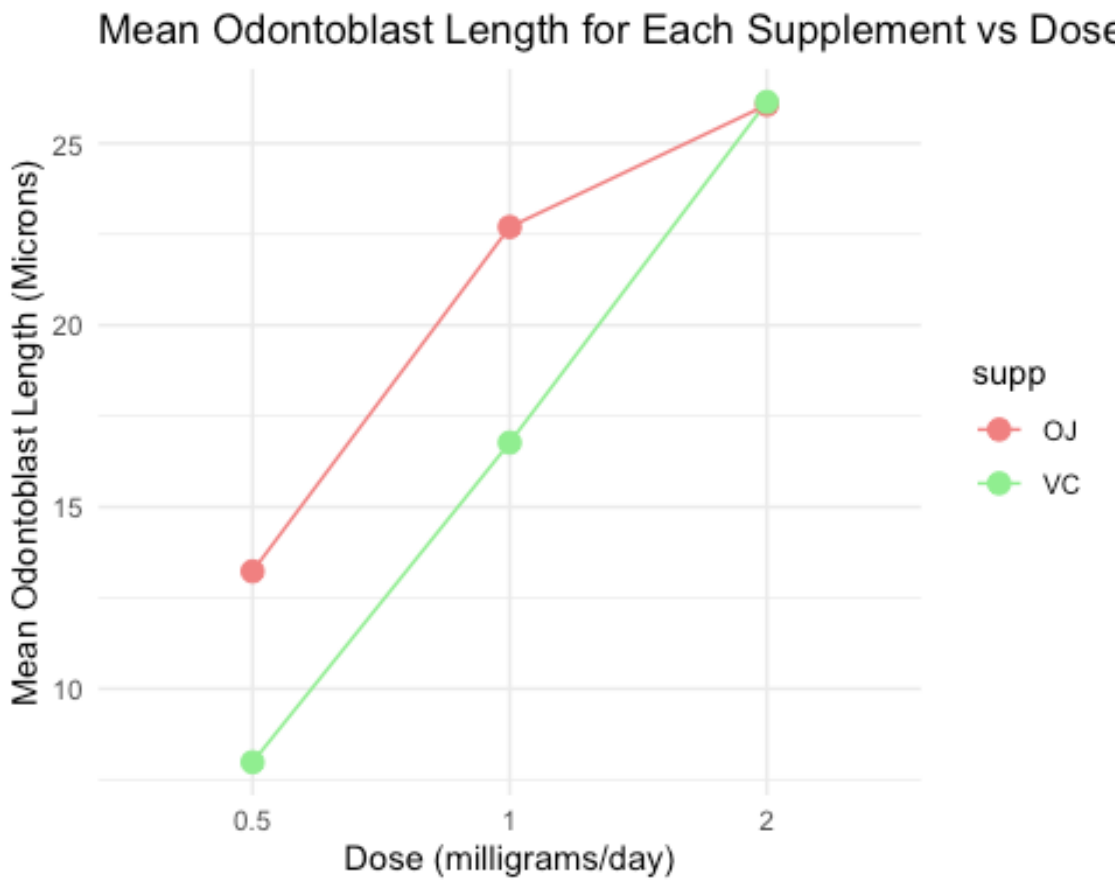


Figure 6. Main Effects Plot for Supplements with Dosage.

Figure 6 shows the means for each supplement, indicating that at a dose of 2.0 mg/day, the means are closely aligned. However, for lower doses, orange juice tends to perform better on average. A crossing of lines indicates that an increase of dosage for one type of supplement has a different effect than it does for the other dosage.

```
# Main Effects Plot for 'supp'
ggplot(df_tooth_growth, aes(x = supp, y = len, group = dose, color = dose)) +
  geom_boxplot() +
  geom_point(position = position_dodge(width = 0.75), size = 2) +
  labs(title = "Main Effects Plot for Supplement", x = "Supplement",
       y = "Tooth Length") +
  scale_color_manual(values = c("lightcoral", "lightgreen", "blue")
                    ) +  # Set custom colors
  theme_minimal()
```

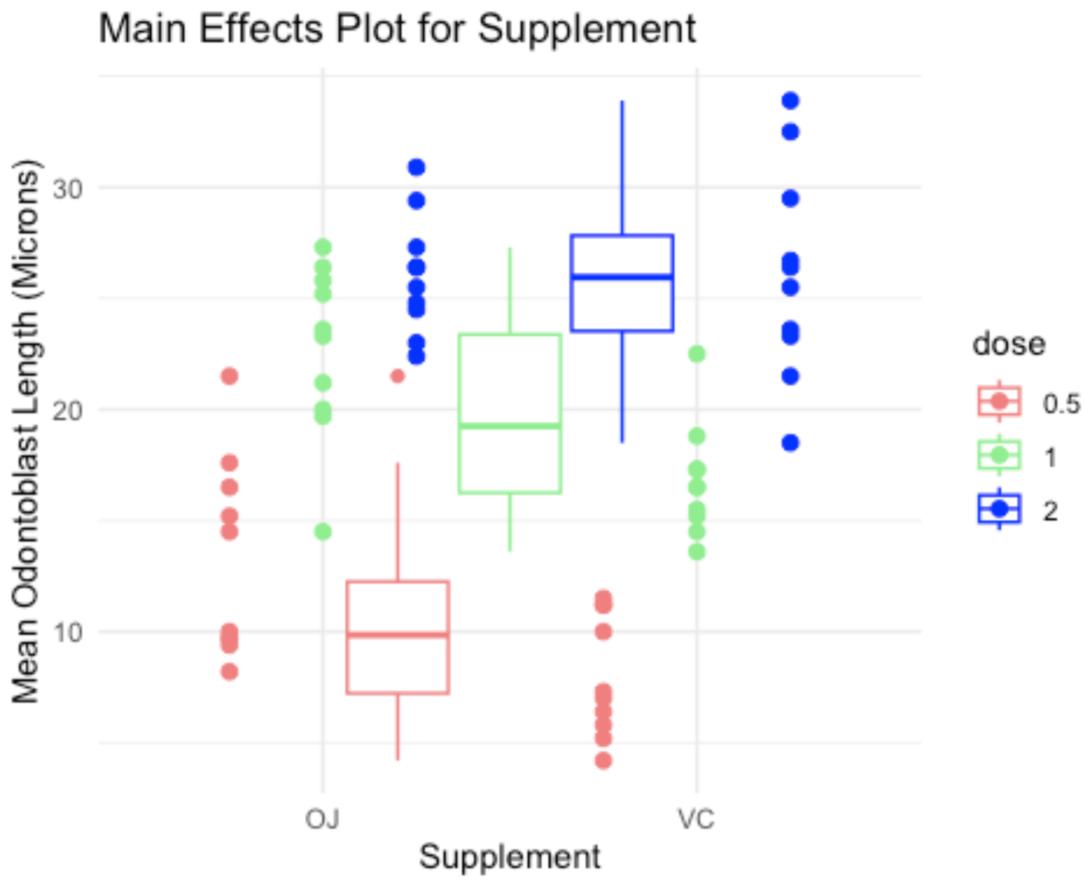


Figure 6. Main Effects Plot for Supplements with Dosage.

From the above plot we can observed that for dosage 0.5 and 1 Orange Juice seems to better improve length of Odontoblast in comparison to Ascorbic Acid. Also, there are a few outliers inducing variability of mean between the groups.

TWO WAY ANOVA

Two-way Analysis of Variance (ANOVA) is a statistical method used to analyze the influence of two categorical independent variables (factors) on a continuous dependent variable. It assesses whether there are significant differences in the means of the dependent variable across the various combinations of levels from both factors. The analysis involves partitioning the total variability in the data into components attributed to each factor, their interaction, and the residual (unexplained) variability. Two-way ANOVA helps identify not only the main effects of each factor but also whether their combined effect is statistically significant.

Null Hypothesis (H0):

1.There is no significant effect of Factor A on the response variable.

2.There is no significant effect of Factor B on the response variable.
3.Null Hypothesis (H_0): There is no interaction between Factor A and Factor B on the response variable.

Alternative Hypothesis (Ha):

1.There is a significant effect of Factor A on the response variable.
2.There is a significant effect of Factor B on the response variable.
3.There is a significant interaction between Factor A and Factor B on the response variable.

Assumptions of Two Way ANOVA

1) Independence.
2) Homogeneity of Variance.
3) Normality.

Hypothesis Testing

1. Shapiro-Wilk normality test to check for Normality:

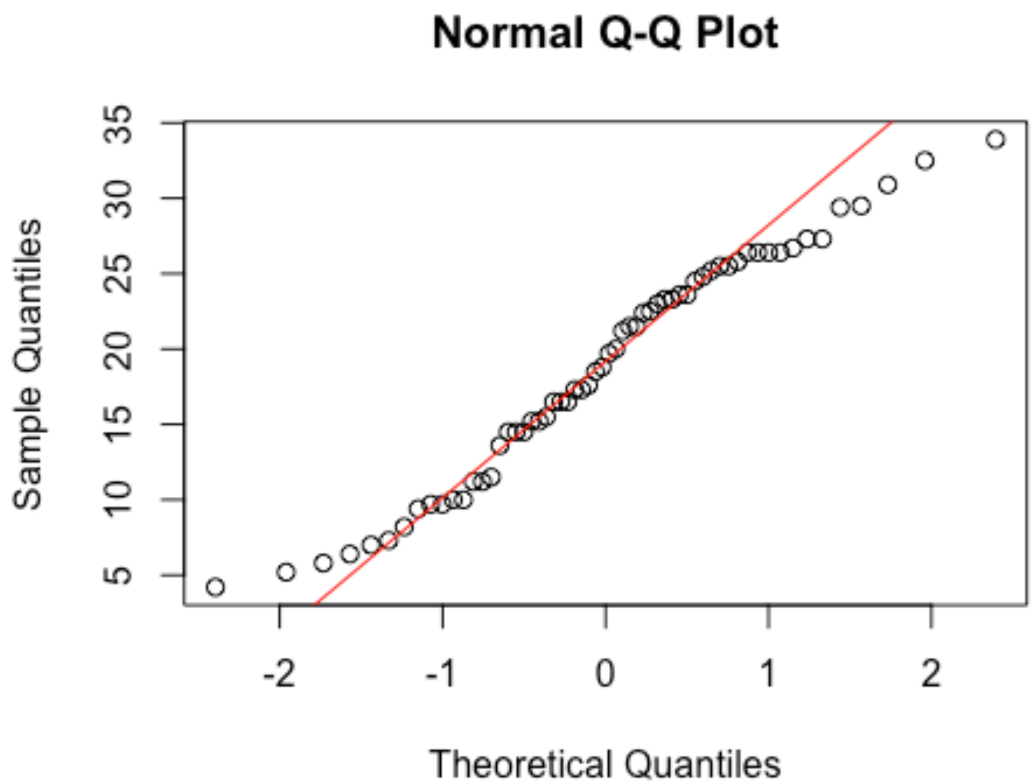
Null Hypothesis (H0):

The data in the sample is normally distributed.

Alternative Hypothesis (Ha):

The data in the sample is not normally distributed.

```
shapiro.test(df_tooth_growth$len)
##
##  Shapiro-Wilk normality test
##
## data:  df_tooth_growth$len
## W = 0.96743, p-value = 0.1091
Since the p-value of 0.1091 is greater than 0.05 significance level,it indica
tes that the data is normally distributed.
qqnorm(df_tooth_growth$len)
qqline(df_tooth_growth$len, col = "red")
```



Since the points are along the straight line except at the ends which denotes the outliers, it indicates that the data is normally distributed for the three treatment groups.

2.Homogeneity of variance Test

```
require(car)
## Loading required package: car
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
library(car)
# Levene test for Interaction Term
#levels(ToothGrowth$supp)
# Convert 'dose' to a factor with specific levels
df_tooth_growth$dose <- factor(df_tooth_growth$dose, levels = c("0.5", "1", "
2"))

# Perform Levene test
leveneTest(len ~ supp * dose, data = df_tooth_growth)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  1.7086 0.1484
##       54
# Levene test for factor1
leveneTest(len ~ supp, data = df_tooth_growth)
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  1.2136 0.2752
##       58
# Levene test for factor2
leveneTest(len ~ dose, data = df_tooth_growth)
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.6457 0.5281
##       57
```

Null Hypothesis (H0):
The null hypothesis for Levene's test is that the variances are equal across all groups.
Alternative Hypothesis (H1):
The alternative hypothesis is that at least one group has a different variance.

Since the p-value is greater than the chosen significance level (0.05), we fail to reject the null hypothesis. The p-value is (0.1484,0.2752, 0.5281), which is greater than 0.05, suggesting evidence against the assumption of difference in variance.

3. Independence

```
# Create a contingency table
contingency_table <- table (df_tooth_growth$supp, df_tooth_growth$dose)

# Conduct Chi-squared test for independence
chi_squared_test <- chisq.test(contingency_table)

# Print the results
print(chi_squared_test)
##
## Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 0, df = 2, p-value = 1
```

From the Chi-Square Test, we see that the p-value of 1 is greater than the significance l evel of 0.05 which denotes that the features are independent.

```
two_way_anova <- aov(len ~ supp * factor(dose), data = df_tooth_growth)
summary(two_way_anova)
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## supp               1  205.4    205.4  15.572 0.000231 ***
```

```
## factor(dose)       2 2426.4   1213.2   92.000  < 2e-16 ***
## supp:factor(dose)  2   108.3     54.2    4.107  0.021860 *
## Residuals         54   712.1    13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Main effects:

- Dose: The F-value for dose is 1213.2 and the p-value is less than 0.0001. This means that there is a statistically significant effect of dose on the residuals. We can conclude that the variability in the residuals is different for different dose levels.
- Supplements: The F-value for supplements is 54.2 and the p-value is 0.0218. This means that there is a statistically significant effect of supplements on the residuals. We can conclude that the variability in the residuals is different for different support levels.

Interaction:

- The F-value for the interaction between dose and supplements is 4.107 and the p-value is 0.0218. This means that there is a statistically significant interaction between dose and supplements. The effect of dose on the residuals is different depending on the level of supplements.

```
TukeyHSD(two_way_anova)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = len ~ supp * factor(dose), data = df_tooth_growth)
##
## $supp
##       diff       lwr       upr     p adj
## VC-OJ -3.7 -5.579828 -1.820172 0.0002312
##
## $`factor(dose)`
##         diff       lwr       upr   p adj
## 1-0.5   9.130  6.362488 11.897512 0.0e+00
## 2-0.5  15.495 12.727488 18.262512 0.0e+00
## 2-1     6.365  3.597488  9.132512 2.7e-06
##
## $`supp:factor(dose)`
##                diff        lwr        upr     p adj
## VC:0.5-OJ:0.5 -5.25 -10.048124 -0.4518762 0.0242521
## OJ:1-OJ:0.5    9.47   4.671876 14.2681238 0.0000046
## VC:1-OJ:0.5    3.54  -1.258124  8.3381238 0.2640208
## OJ:2-OJ:0.5   12.83   8.031876 17.6281238 0.0000000
## VC:2-OJ:0.5   12.91   8.111876 17.7081238 0.0000000
```

```
## OJ:1-VC:0.5   14.72   9.921876 19.5181238 0.0000000
## VC:1-VC:0.5    8.79   3.991876 13.5881238 0.0000210
## OJ:2-VC:0.5   18.08  13.281876 22.8781238 0.0000000
## VC:2-VC:0.5   18.16  13.361876 22.9581238 0.0000000
## VC:1-OJ:1     -5.93 -10.728124 -1.1318762 0.0073930
## OJ:2-OJ:1      3.36  -1.438124  8.1581238 0.3187361
## VC:2-OJ:1      3.44  -1.358124  8.2381238 0.2936430
## OJ:2-VC:1      9.29   4.491876 14.0881238 0.0000069
## VC:2-VC:1      9.37   4.571876 14.1681238 0.0000058
## VC:2-OJ:2      0.08  -4.718124  4.8781238 1.0000000
```

The **VC** group has a significantly lower tooth length compared to the **OJ** group, with a difference of -3.7 units (p-value = 0.0002312).

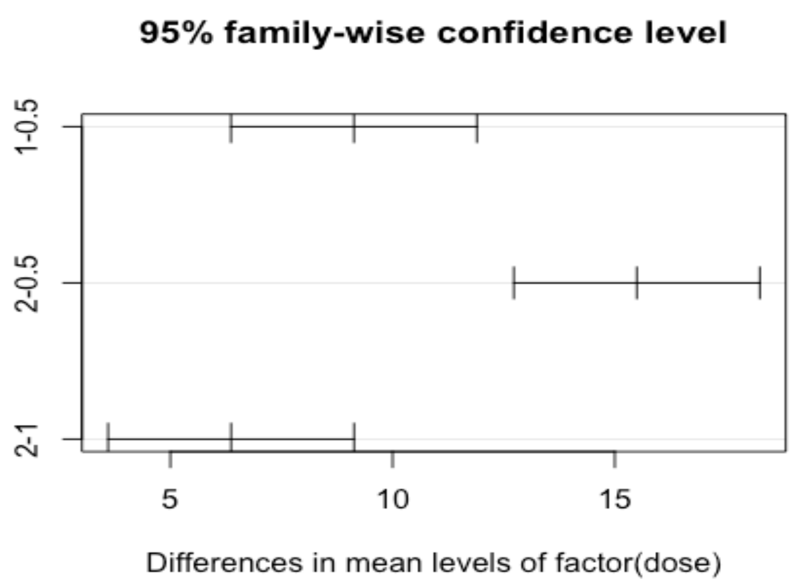
The tooth length increases significantly with a dose of 1 compared to 0.5 (p-value = 0.0e+00).
The tooth length increases significantly with a dose of 2 compared to 0.5. (p-value = 0.0e+00)
The tooth length increases significantly with a dose of 2 compared to 1 (p-value = 2.7e-06).

We can see Significant Differences in each of the interaction comparisons except for the differences between the below:

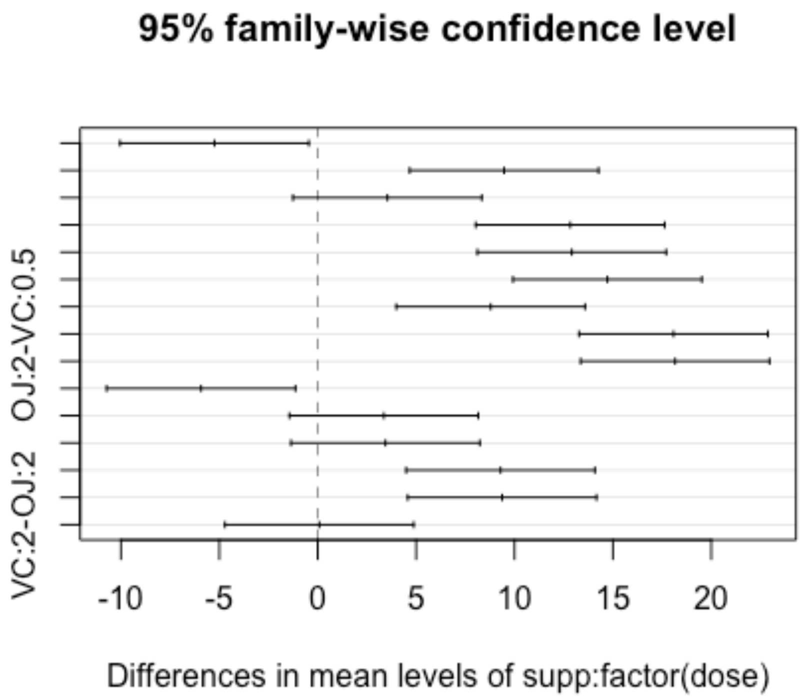
(VC:1-OJ:0.5, OJ:2-OJ:1, VC:2-OJ:1, VC:2-OJ:2)

```r
plot(TukeyHSD(two_way_anova))
```

The above figure shows the mean difference of -3.7 with confidence intervals from -5.5 to -1.8.



The above figure shows the mean difference of 9, 15 and 6 with confidence intervals ranging from 3 to 18.



The above figure shows different mean differences with confidence intervals ranging from -10 to 22.

```
require(rstatix)
library(rstatix)

eta_squared(two_way_anova)
##              supp      factor(dose) supp:factor(dose)
##        0.05948365        0.70286419        0.03137672
partial_eta_squared(two_way_anova)
##              supp      factor(dose) supp:factor(dose)
##         0.2238254         0.7731092         0.1320279
```

Eta Squared:

- This indicates that approximately 5.95% of the total variability in the response variable (e.g., tooth length) can be attributed to the effect of the supplement type (Supplements).
- A large value of 70.29% suggests that the factor "dose" explains a substantial portion of the total variability in the response variable.
- About 3.14% of the total variability is attributed to the interaction between supplement type and dose.

Partial Eta Squared:

- This 0.2238 proportion of variability in the response variable is uniquely explained by the supplement type, considering the effect of other factors.
- Similarly, this 0.7731 proportion of variability in the response variable is uniquely explained by the dose factor.
- 0.1320 proportion of variability uniquely explained by the interaction between supplement type and dose.

CONCLUSIONS

1) There is significant difference in means between the groups of factor A(Supplements).
2) There is significant difference in means between the groups of factor B(Dose).
3) We can observe that there is interaction between the two factors (Supplement and Dose).

REFERENCES

1) CRAMPTON EW. The growth of the odontoblasts of the incisor tooth as a criterion of the vitamin C intake of the guinea pig. J Nutr. 1947 May;33(5):491-504. doi: 10.1093/jn/33.5.491. PMID: 20294344.
2) Bliss C.I. (1952). *The Statistics of Bioassay*. Academic Press.
3) https://www.stat.cmu.edu/~hseltman/309/Book/chapter11.pdf