Analyzing the ToothGrowth data

Abhijeet

1. Synopsis

This project analyses the Tooth Growth data in the R datasets package and tries to find the correlation between the dose of supplement and tooth length in guinea pigs.

1.1. Loading Data

```
library(datasets)
library(ggplot2)

data<-ToothGrowth
head(data)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

1.2. Data Summary

```
summary(data)
```

```
##       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
```

1.3. Exploring the data

```
# Count any missing values
sum(is.na(data))
```

```
## [1] 0
```

Finding all unique values

```
unique(data$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(data$dose)
```

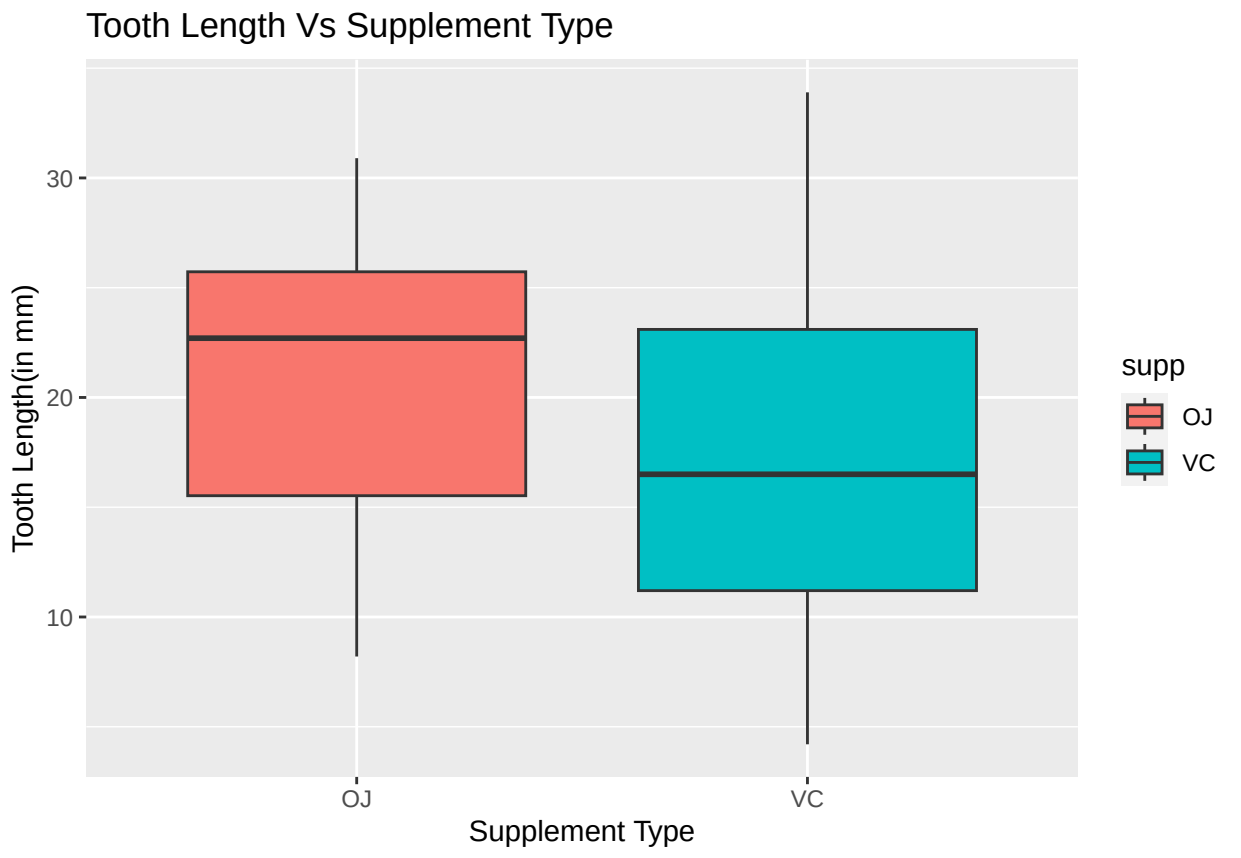
```
## [1] 0.5 1.0 2.0
```

Checking How the suppliment type and Dose are distributed in the datasets.

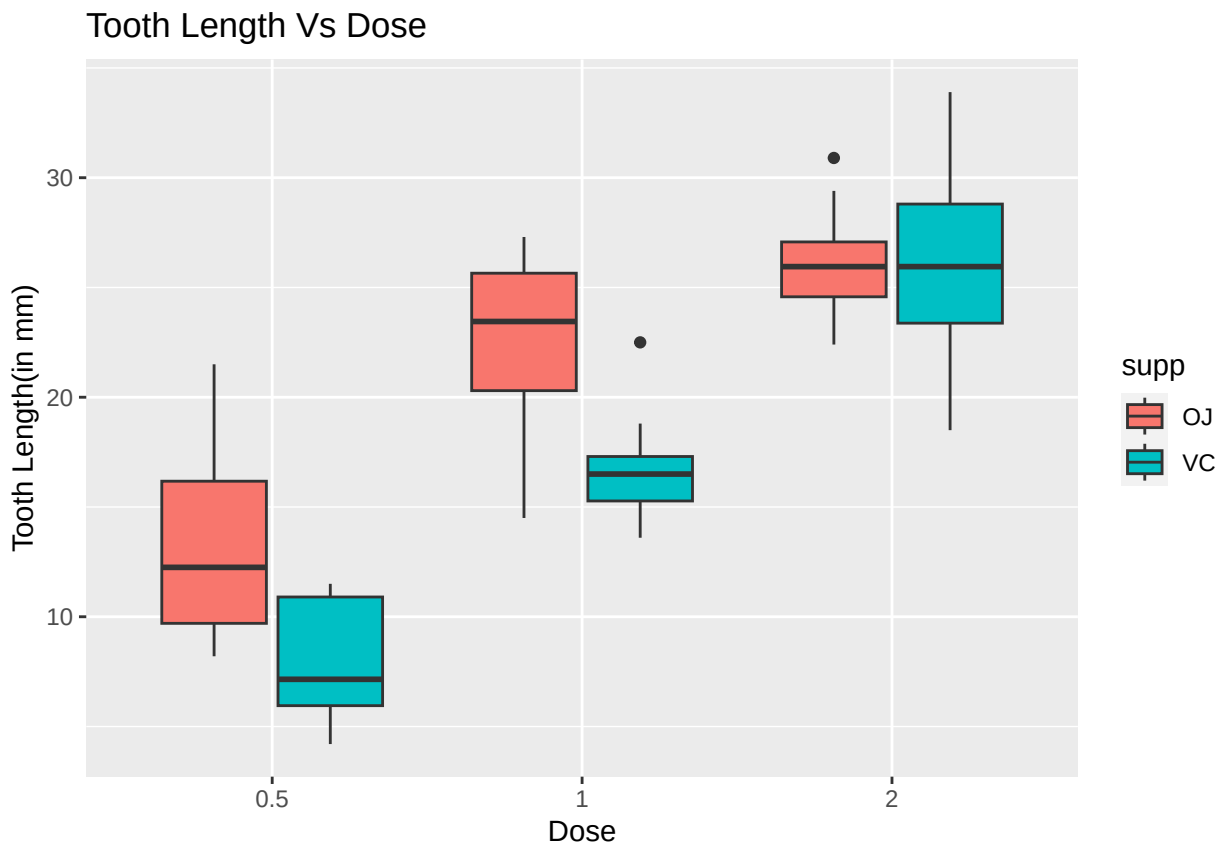
```
table(data$dose,data$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
ggplot(data,aes(supp,len))+geom_boxplot(aes(fill=supp))+xlab('Supplement Type')+
  ylab('Tooth Length(in mm)')+labs(title="Tooth Length Vs Supplement Type")
```

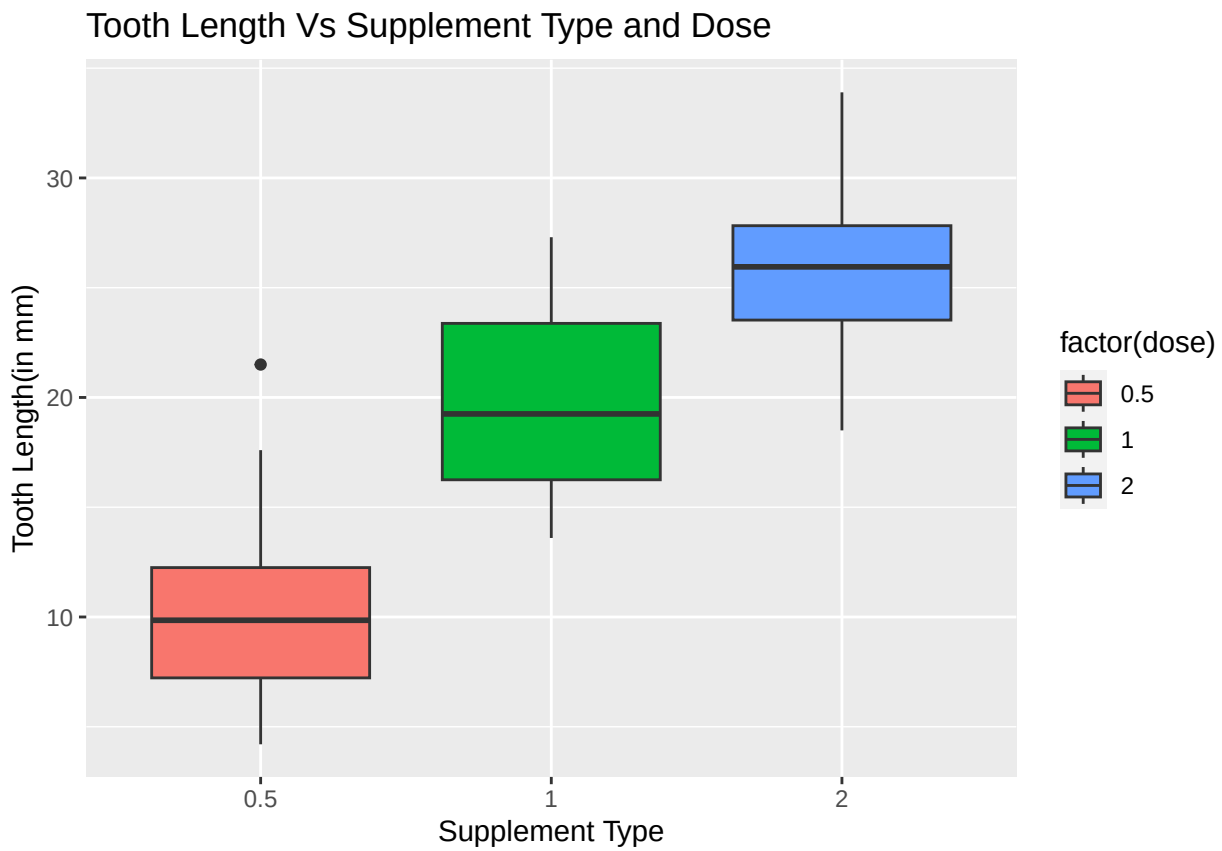


```
ggplot(data,aes(factor(dose),len))+geom_boxplot(aes(fill=supp))+xlab('Dose')+ylab('Tooth Length(in mm)')
```



The tooth length for the specific dose amount of Orange Juice is higher than the Vitamin C dose amount.

```
ggplot(data,aes(factor(dose),len))+geom_boxplot(aes(fill=factor(dose)))+xlab('Supplement Type')+
  ylab('Tooth Length(in mm)')+labs(title="Tooth Length Vs Supplement Type and Dose")
```



Tooth Length Varies as the amount of dose is varying. ### 2. Hypothesis Testing ### 2.1. Hypothesis 1 Is Tooth Length depends on the amount/dose of supplements Vitamin C?

```
dose_VC_1<- subset(data, supp == 'VC' & dose %in% c(0.5, 1.))
dose_VC_2<- subset(data, supp == 'VC' & dose %in% c(0.5, 2.))
dose_VC_3<- subset(data, supp == 'VC' & dose %in% c(1., 2.))

t.test(len~dose, data=dose_VC_1, var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean in group 0.5   mean in group 1
##              7.98             16.77
```

```
t.test(len~dose, data=dose_VC_2, var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  len by dose
## t = -10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -21.90151 -14.41849
## sample estimates:
## mean in group 0.5   mean in group 2
##              7.98             26.14
```

```
t.test(len~dose, data=dose_VC_3, var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean in group 1 mean in group 2
##           16.77           26.14
```

For rejecting the null hypothesis p-value should be less than 0.05 which is the 5% significance level. The p-value for all three tests is less then 0.05. Hence, the null hypothesis can be rejected and it suggests that there is a correlation exists between tooth length and supplement amount of "Vitamin C".

2.2. Hypothesis 2

Is Tooth Length depends on the amount/dose of supplements Orange Juice?

```
dose_OJ_1<- subset(data, supp == 'OJ' & dose %in% c(0.5, 1.0))
dose_OJ_2<- subset(data, supp == 'OJ' & dose %in% c(0.5, 2.))
dose_OJ_3<- subset(data, supp == 'OJ' & dose %in% c(1., 2.))

t.test(len~dose, data=dose_OJ_1, var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean in group 0.5   mean in group 1
##             13.23             22.70
```

```r
t.test(len~dose, data=dose_OJ_2, var.equal=FALSE, paired=FALSE)
```

```
##
##  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
```

```r
t.test(len~dose, data=dose_OJ_3, var.equal=FALSE, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

Similarly, The P-values in the Hypothesis 2's tests is less than 0.05. Hence, the null hypothesis can be rejected and it suggests that there is a correlation exists between tooth length and supplement amount of "Orange Juice".

2.3. Hypothesis 3

Is Tooth Length depends on just supplement Type?

```r
df<-data
t.test(len~supp, data=df, var.equal=FALSE, paired=FALSE)
```

```
##
##  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 between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

P-value should be less than 0.05 for rejecting Null, which is not true in this test. Hence, the null hypothesis that the tooth length depends on just supplement type cannot be rejected.

2.4. Hypothesis 4

is Tooth Length depends on specific dose amount?

```
dose_5<-subset(data, dose %in% c(0.5))
dose_1<-subset(data, dose %in% c(1.))
dose_2<-subset(data, dose %in% c(2.))
t.test(len~supp,data=dose_5,var.equals=FALSE,paired=FALSE)
```

```
##
##  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
```

```
t.test(len~supp,data=dose_1,var.equals=FALSE,paired=FALSE)
```

```
##
##  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
```

```
t.test(len~supp,data=dose_2,var.equals=FALSE,paired=FALSE)
```

```
##
##  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
```

The P-value for dose amount of 0.5mg & 1mg can be rejected. Hence, the tooth length can vary with these dose levels.However, The P-value for 2.0mg is 0.9639 that means the null hypothesis can not be rejected.

3. Conclusion

To make conclusions it is assumed that the data taken from guinea pigs is independent, random sample, continuous and normally distributed. It appears that the tooth length is higher for orange juice than that of only Vitamin C, and the dose level has significant level of impact o tooth level for 0.5mg and 1mg, but not for 2mg. Based on the evidence, We can confirm that the orange juice is a better delivery method for the dose of Vitamin C, but at the higher dose level there is no improvements.

==========