Data Analysis of the ToothGrowth Dataset

1. Synopsis

The ToothGrowth data in the R datasets package contains information about the effect of Vitamin C On tooth growth in guinea pigs. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

In this project, we are going to analyze the ToothGrowth data and perform some basic inferential data analysis.

2. Load the ToothGrowth data and perform some basic exploratory data analyses. Provide a basic summary of the data.

Load in the required libraries.

```
library(ggplot2)
```

We load the dataset, observe the first few rows and convert the `dose` variable from numeric to categorical. We then observe the structure of the dataset to ensure that the change was made.

```
data(ToothGrowth)
head(ToothGrowth)
```

```
##    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
```

```
# Convert dose to a factor variable
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
str(ToothGrowth)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

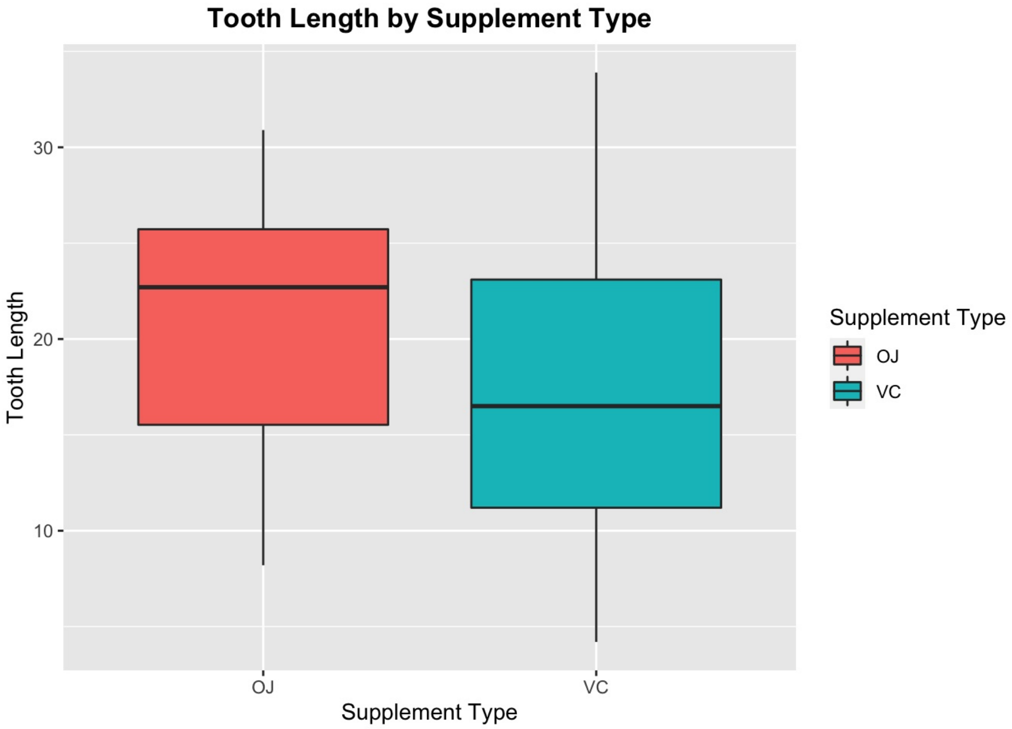
We observe a summary of the data as follows:

```
summary(ToothGrowth)
```

```
##       len        supp     dose   
##  Min.   : 4.20   OJ:30   0.5:20  
##  1st Qu.:13.07   VC:30   1  :20  
##  Median :19.25           2  :20  
##  Mean   :18.81                   
##  3rd Qu.:25.27                   
##  Max.   :33.90                   
```

Let’s plot a boxplot of tooth length against supplement type.

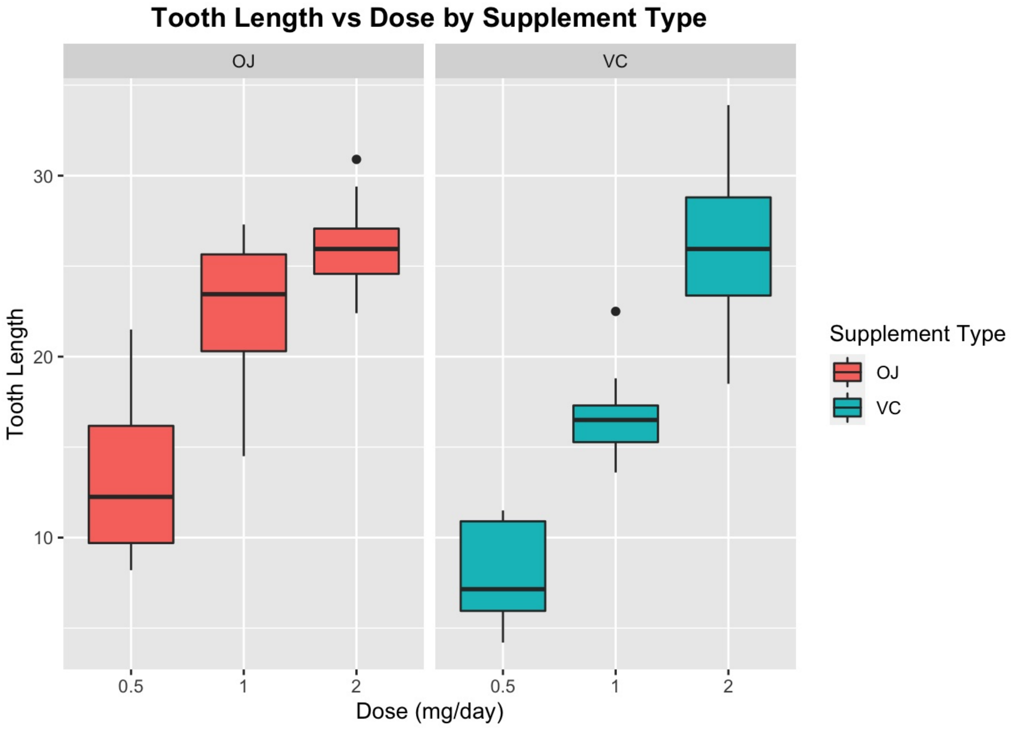
```
ggplot(ToothGrowth, aes(x=supp,y=len, fill=supp)) + geom_boxplot() + xlab("Supplement Type") + ylab("Tooth Length"
) + ggtitle("Tooth Length by Supplement Type") + theme(plot.title = element_text(hjust = 0.5, face="bold")) + guid
es(fill = guide_legend(title = "Supplement Type"))
```



We observe that the median tooth lengths for those supplemented by orange juice is larger than that for those supplemented by ascorbic acid, indicating that orange juice appears to be more effective in promoting tooth growth.

To have a better understanding, we plot a boxplot of tooth length against dose for the different supplements.

```
ggplot(ToothGrowth, aes(x=dose,y=len, fill=supp)) + geom_boxplot() + facet_grid(.~supp) + xlab("Dose (mg/day)") +
ylab("Tooth Length") + ggtitle("Tooth Length vs Dose by Supplement Type") + theme(plot.title = element_text(hjust
= 0.5, face="bold")) + guides(fill = guide_legend(title = "Supplement Type"))
```



We observe that in general, larger dosages of Vitamin C result in longer tooth lengths. For 0.5 and 1 mg/day dosages, the tooth lengths for those supplemented with orange juice is greater than those supplemented by ascorbic acid. For 2 mg/day dosage, the median tooth lengths for both supplement types appear similar but the variance for those supplemented with ascorbic acid appears to be much larger.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. State your conclusions and the assumptions needed for your conclusions.

3.1 Does the difference in supplement types affect tooth growth?

Assuming that the variance in tooth length for different supplement types are different, we can perform a two sample t-test to determine if the mean in tooth length for different supplement types are equal. The hypothesis test is as follows:

- H_0 : Means are equal, i.e. Difference in means is 0
- H_1 : Means are not equal, i.e. Difference in means is not 0

```
t.test(len ~ supp, var.equal = FALSE, data=ToothGrowth)
```

```
##
##  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 OJ mean in group VC
##         20.66333         16.96333
```

We observe that the p-value = 0.06 > 0.05 (critical value) and the 95% confidence interval is (-0.1710156, 7.5710156), which contains 0. This indicates that we are unable to reject H_0 at the 95% confidence level, where the significance level, α = 0.05. We are unable to conclude that the means in tooth length for different supplement types are not equal, indicating that the difference in supplement types may not affect tooth growth.

3.2 Does the difference in dosages affect tooth growth?

Assuming that the variance in tooth length for different dosages are different, we can also perform pairwise t-tests to determine if the mean in tooth length for different dosages are equal.

We first subset the data into the different dosages and extract out the `len` column.

```
low_dose <- ToothGrowth[ToothGrowth$dose == 0.5,1]
med_dose <- ToothGrowth[ToothGrowth$dose == 1,1]
high_dose <- ToothGrowth[ToothGrowth$dose == 2,1]
```

Now, we perform pairwise t-tests between the different dosages. Similarly, the hypothesis test is as follows:

- H_0 : Means are equal, i.e. Difference in means is 0
- H_1 : Means are not equal, i.e. Difference in means is not 0

```
t.test(low_dose, med_dose, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  low_dose and med_dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(med_dose, high_dose, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  med_dose and high_dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

```
t.test(low_dose, high_dose, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  low_dose and high_dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

We observe that all of their p-values are less than the critical value of 0.05 and that all of the 95% confidence intervals do not contain 0. This indicates that we are able to reject H_0 at the 95% confidence level, suggesting that the mean in tooth length for different dosages are all different. We can conclude that the difference in dosages does affect tooth growth.