Part 2 - Inferential Analysis using *ToothGrowth* data

Overview

Analyze the ToothGrowth data in the R datasets package.

Instructions

1. Load the ToothGrowth data and perform some basic exploratory data analysis
2. Provide a basic summary of the data
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose
4. State your conclusions and the assumptions needed for your conclusions

Prepare Environment

Load Libraries and set Global Options

```
#To suppress loading messages set *message = FALSE*.

#Set global options *echo = TRUE* so others will be able to read the code a
nd set *results = hold* to hold & push output to end of chunk.

library(knitr)

opts_chunk$set(echo = TRUE, results = 'hold')

library(data.table)

library(ggplot2)
```

1. Load the ToothGrowth dataset

```
toothGrowth <- data.table(ToothGrowth)
```

Description 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).

Format A data frame with 60 observations on 3 variables.

1a. Exploratory Data Analysis

```
head(toothGrowth); tail(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
##     len supp dose
## 1: 24.8   OJ    2
## 2: 30.9   OJ    2
## 3: 26.4   OJ    2
## 4: 27.3   OJ    2
## 5: 29.4   OJ    2
## 6: 23.0   OJ    2
```

The three variables are **length**, **supplement**, and **dose**.

```
summary(ToothGrowth)
```

```
##       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
```

```
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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

There are two levels (*delivery types*) for **supp** (supplement): **OJ** (Orange Juice) and **VC** (Vitamin C).

Length and **dose** are both numeric values but we cannot determine how many values exist for **dose**. View the unique values of **dose**.

```
unique(ToothGrowth$dose)
```

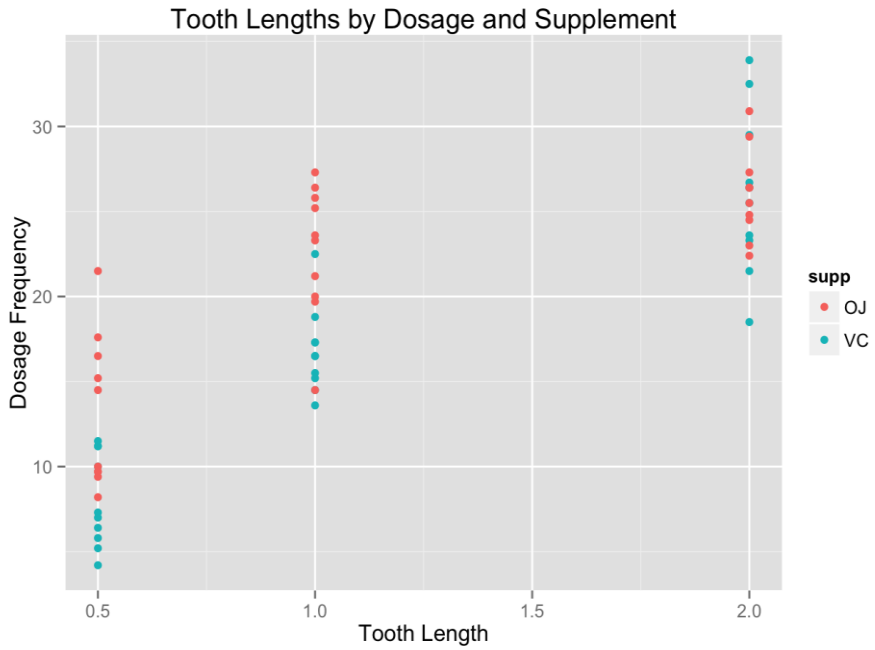
```
## [1] 0.5 1.0 2.0
```

There are three discrete levels for **dose**: *0.5*, *1.0*, and *2*.

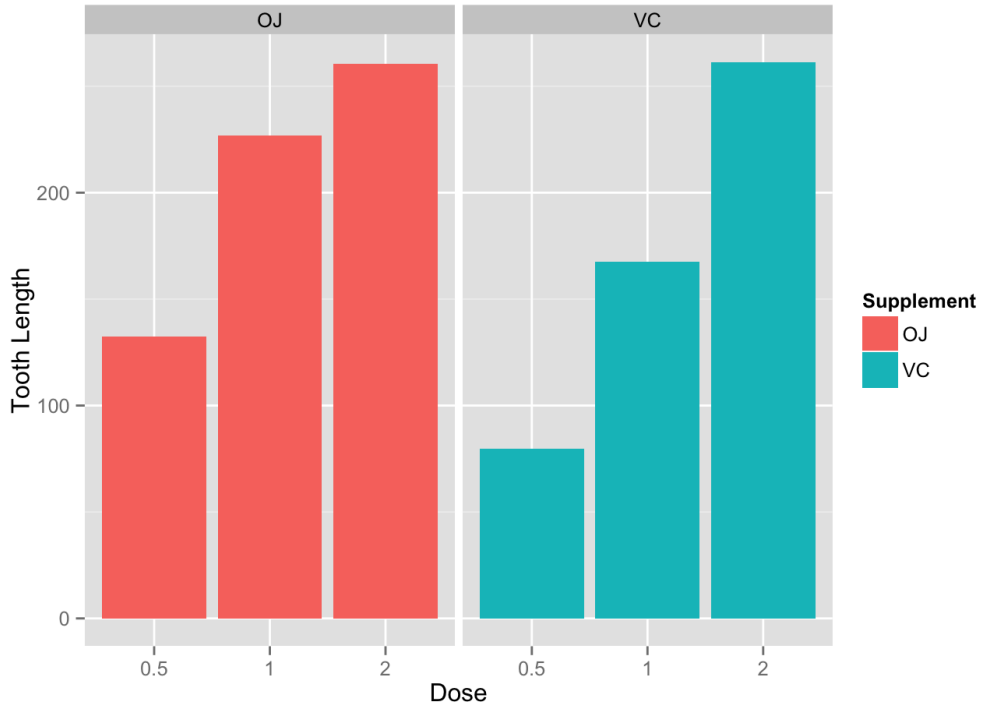
Visually examine the data by looking at the tooth **length** compared to **dose** by **supplement**.

```
ggplot(aes(x=dose, y = len), data = ToothGrowth) +
```

```
     labs(x = "Tooth Length", y = "Dosage Frequency", title = "Tooth Length
s by Dosage and Supplement") +
    geom_point(aes(color = supp))
```




```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity") +
    facet_grid(. ~ supp) +
    xlab("Dose") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="Supplement"))
```



2. Summary of initial data analysis

There appears to be an impact on tooth growth by increasing the dosage but it is unclear which of the two supplements contributes to the growth.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Use *t.test* to determine if there is a difference in the performance of the treatments. First, we will run the test based on **supplement**. Looking to see if the **p-value > 0.05** and if the **confidence interval** crosses **0**.

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"], ToothGrowth$len[ToothGrowth
$supp=="VC"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[Too
thGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

The **p-value** is **0.06063** and the **confidence interval** is [**-0.1710156 7.5710156**], thus containing **0**.

Since the **p-value** is **0.06063**, there is not enough evidence to reject the null hypothesis. We cannot assume the delivery type has a significant effect on tooth growth.

Test the tooth length comparing the dosage of 1mg to 2mg to determine the effects of an increased dosage.

```
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

We see the **p-value** is very small, and is significant. Therefore, we can reject the null hypothesis and assume the dosage increase from 1mg to 2mg creates an positive effect on tooth growth.

Next, perform the test comparing the dosage of 0.5mg to 1mg.

```
t.test(ToothGrowth$len[ToothGrowth$dose==1], ToothGrowth$len[ToothGrowth$dose==0.5], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
```

```
##    19.735    10.605
```

Again, we see the **p-value** is still small although slighty larger than the previous test, therefore, it is significant.

We can again reject the null hypothesis and assume the dosage increase from .5mg to 1mg creates an positive effect on tooth growth.

There is no need for futher testing of dosages given the previous tests.

4. State conclusions and the assumptions needed for conclusions

In this experiment, we assume there is a common variance in the population and that the guinea pigs were chosen at random.

The delivery type does not show a significant increase in tooth growth even though it does have a confidence level that crosses 0 at the 95% confidence.

However, there does appear to be a difference with an increase in tooth growth when the dosage is increased. The tests comparing the dosage show confidence intervals of differences never crossing zero.

Conclusion

Increasing the dosage leads to an increase in tooth growth in guinea pigs.