Part 2

Ayush Kumar

20/10/2020

Basic Inferential Data Analysis-We are going to analyze the ToothGrowth data in the R datasets package.

##Q1 Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(ggplot2)
data(ToothGrowth)
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 ...
```

```
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
```

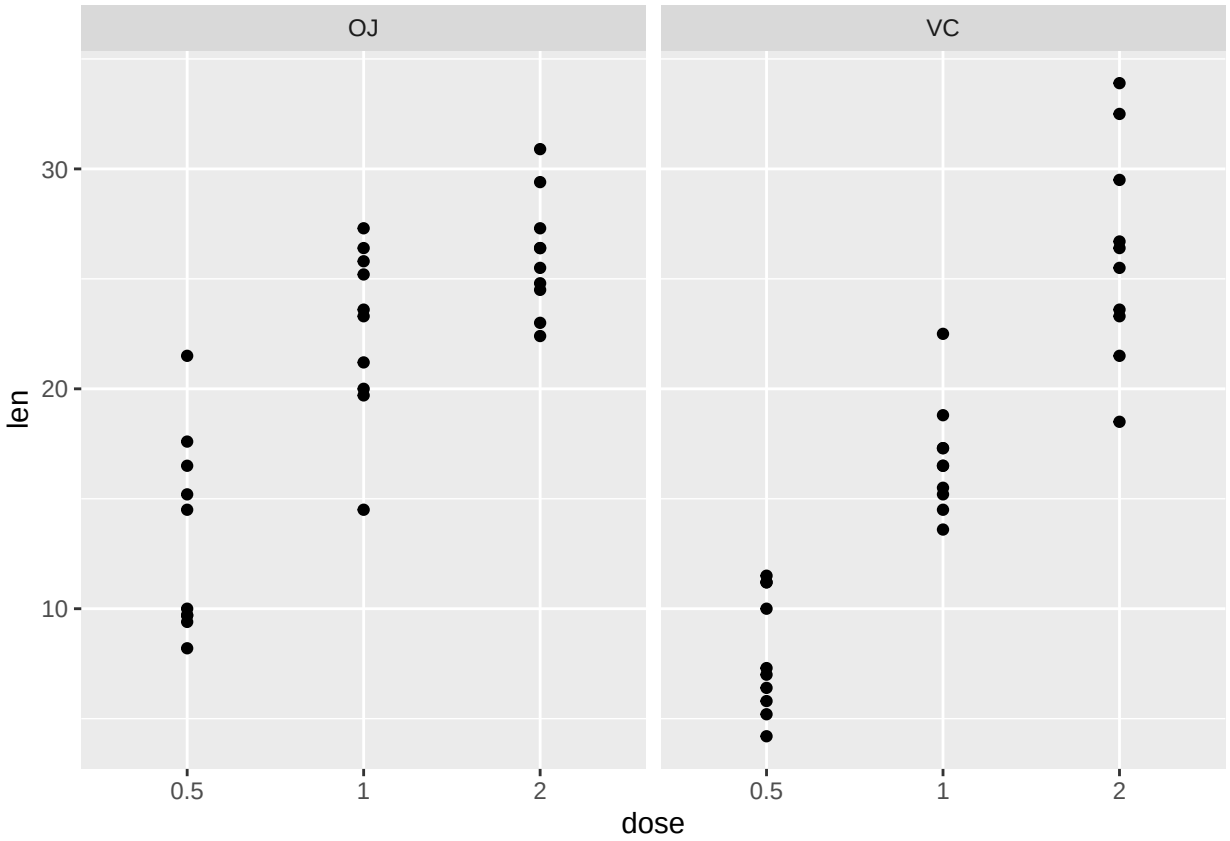
```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
str(ToothGrowth)
```

```
## '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 ...
```

```
qplot(dose, len, data=ToothGrowth, facets=.~supp, geom=c("point", "smooth"), method="loess")
```

```
## Warning: Ignoring unknown parameters: method
```

```
## `geom_smooth()` using formula 'y ~ x'
```

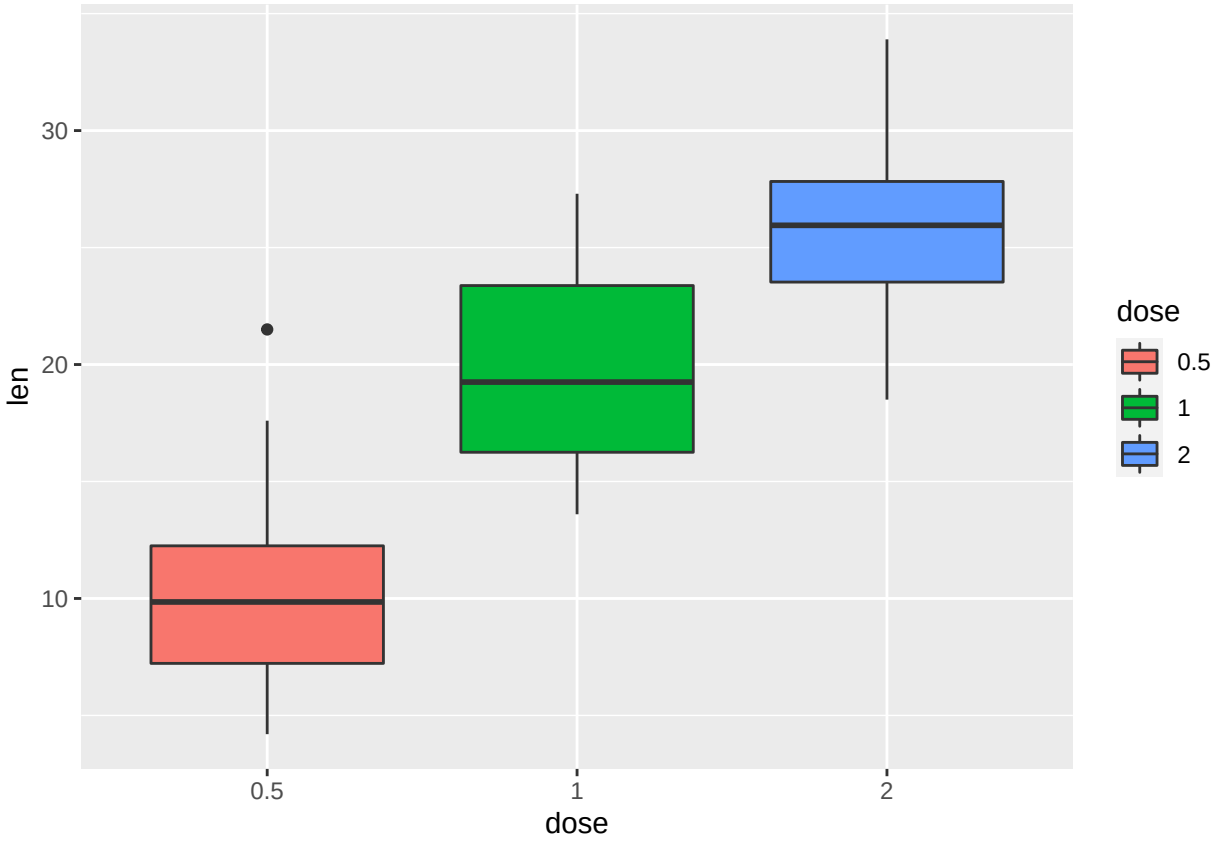


##Q2 Provide a basic summary of the data.

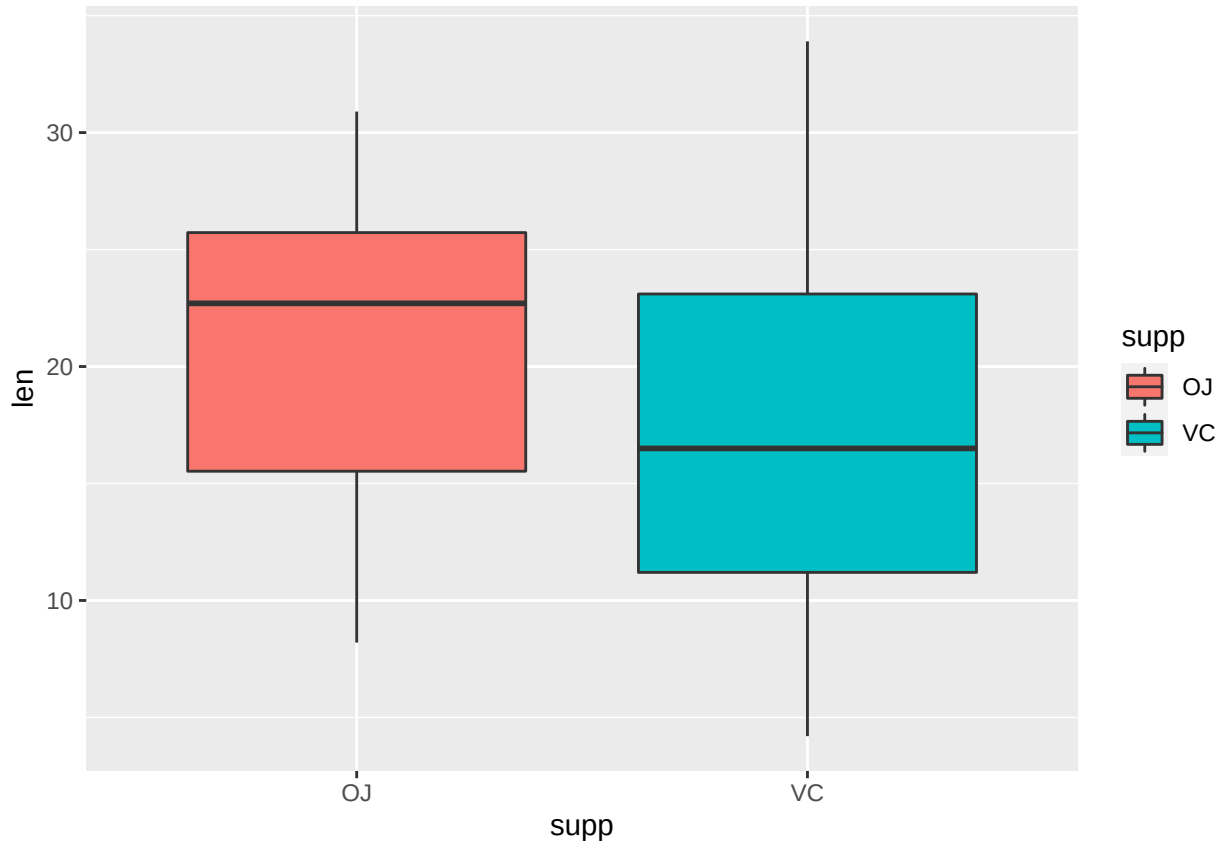
```
summary(ToothGrowth)
```

```
##       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
```

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




```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```



##Q3 Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

```
varsupp <- ToothGrowth$supp
vardose <- ToothGrowth$dose
varlen <-  ToothGrowth$len
```

```
t.test(varlen[varsupp == "VC"],varlen[varsupp == "OJ"], paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  varlen[varsupp == "VC"] and varlen[varsupp == "OJ"]
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

```
t.test(varlen[vardose == 0.5],varlen[vardose == 1], paired=FALSE)
```

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

```
## data:  varlen[vardose == 0.5] and varlen[vardose == 1]
## 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
```

##Q4 State your conclusions and the assumptions needed for your conclusions.

Conclusion-

1. For guiena pigs Vitamin c is a factor of growth for their teeth.
2. Whereas the delivery mode does not have any imapact on their growth.

Assumption-

1.We assume that variance in all groups should be expected to be equal. 2.The underlying assumption is that sampling of Guinea Pigs to assign them to a supplement and a dose was done properly.