# Data Science Course with R:Statistical Inference:Course Project

Question 2:Basic Inferential Data Analysis Instructions

Sagnik Das

6/24/2020

**Load the ToothGrowth data and perform some basic exploratory data analyses**

**Answer:**

```
library(datasets)
library(ggplot2)

dat<- as.data.frame(ToothGrowth)

head(dat)
```

```
##    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
```

```
names(dat)
```

```
## [1] "len"  "supp" "dose"
```

```
dat$dose<- as.factor(dat$dose)

str(dat)
```

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

```
summary(dat)
```

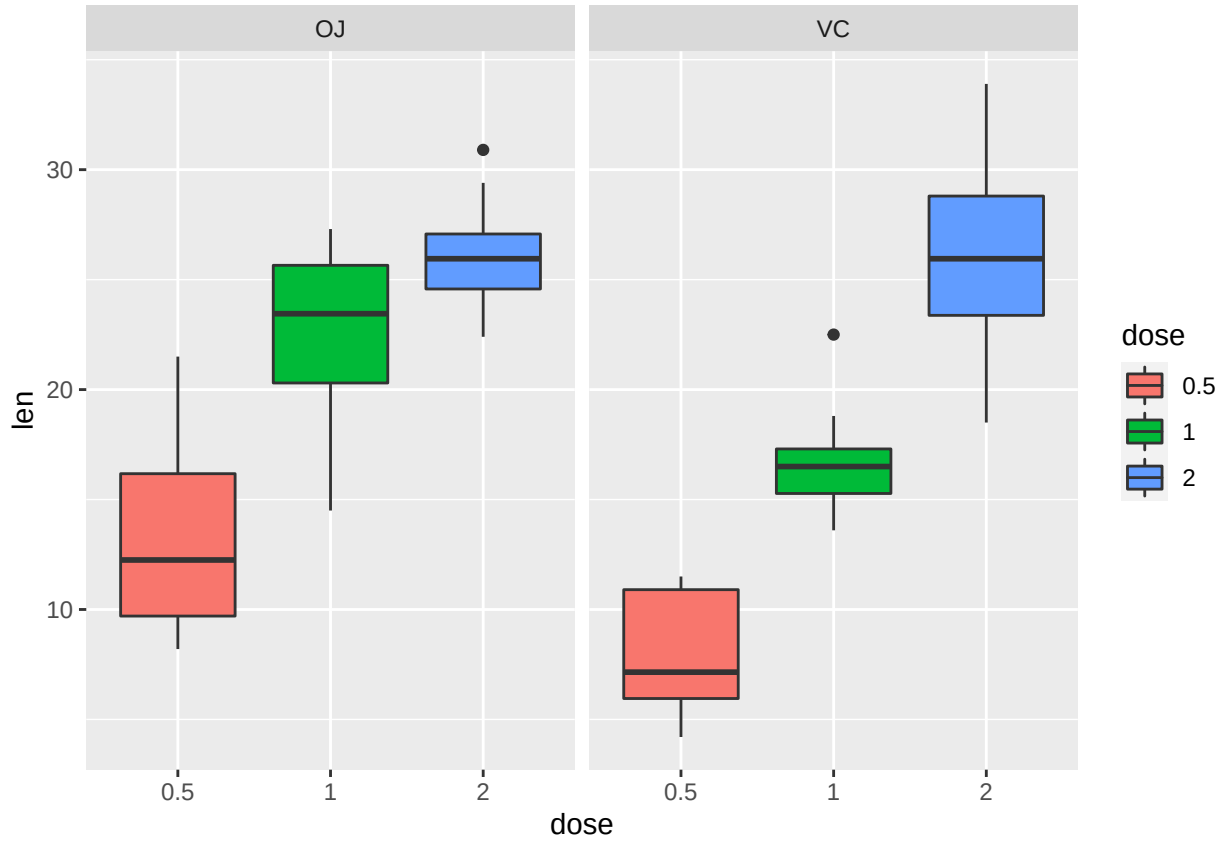
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
```

```
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

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

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

```
p1<- ggplot(dat, aes(x=dose,y=len,fill=dose)) + geom_boxplot() + theme(legend.position = "right") + face
p1
```




The above set of results provide a brief summary of the dataset which primarily consists of 3 variables: The length of teeth of guinea pig given by the variable "len", the suppliment given to the guniea pigs represented by tghe variable "supp" and lastly the variable "dose" which is a factor variable which takes 3 values and represents the dosages of the suppliments given to the guinea pigs. The above figures records the growth in the in the tooth length of the guinea pigs, according to the dosage amounts and what kind of supplement they were given.

## Provide a basic summary of the data.

The first figure shows that increasing the dosage of the supplements increases the size of the tooth of the guinea pigs. It also shows that for dosage amounts 0.5 and 1, Orange Juice is a more effective supplement compared to Ascorbic Acid when it comes to helping tooth growth. However, when the dosage amounts to 2 miligrams, the difference in the effect of the 2 supplements are not that distinct. Also, for a particular supplement, the difference in the effect for different dosage amounts are much more profound in case of ascorbic acid than orange juice.

## Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

```
hyp1<- t.test(dat$len~dat$supp)

print(c(paste("Confidence Interval:",hyp1$conf.int), paste("P-Value:",hyp1$p.value )))
```

**Hypothesis 1: Hypothesis to check if there are statistically significant difference in the mean growth of length of tooth according to the difference in the supplements.**

```
## [1] "Confidence Interval: -0.171015618367164"
## [2] "Confidence Interval: 7.57101561836716"
## [3] "P-Value: 0.0606345078809341"
```

Interpretation: We fail to reject the null hypothesis that there is no statistically significant difference between the effect of the supplements across all dosages. The p-value is greater than 5% which is the level of significance.

```
hyp2<- t.test(len~supp,data = subset(dat,dose==0.5))

print(c(paste("Confidence Interval:",hyp2$conf.int), paste("P-Value:",hyp2$p.value )))
```

**Hypothesis 2: Hypothesis to check if there are statistically significant difference in the mean growth of length of tooth according to the difference in the supplements when dose==0.5**

```
## [1] "Confidence Interval: 1.71905727146767"
## [2] "Confidence Interval: 8.78094272853233"
## [3] "P-Value: 0.0063586067640968"
```

Interpretation: We reject the null hypothesis that there is no statistically significant difference between the effect of the supplements across all dosages. The p-value is less than 5% which is the level of significance. Thus we conclude statistically significant difference in the effects of the supplements when dosage is 0.5.

```r
hyp3<- t.test(len~supp,data = subset(dat,dose==1))

print(c(paste("Confidence Interval:",hyp3$conf.int), paste("P-Value:",hyp3$p.value )))
```

**Hypothesis 3: Hypothesis to check if there are statistically significant difference in the mean growth of length of tooth according to the difference in the supplements when dose==1**

```
## [1] "Confidence Interval: 2.80214824916537"
## [2] "Confidence Interval: 9.05785175083463"
## [3] "P-Value: 0.00103837587229988"
```

Interpretation: We reject the null hypothesis that there is no statistically significant difference between the effect of the supplements across all dosages. The p-value is less than 5% which is the level of significance. Thus we conclude statistically significant difference in the effects of the supplements when dosage is 1.

```r
hyp4<- t.test(len~supp,data = subset(dat,dose==2))

print(c(paste("Confidence Interval:",hyp4$conf.int), paste("P-Value:",hyp4$p.value )))
```

**Hypothesis 4: Hypothesis to check if there are statistically significant difference in the mean growth of length of tooth according to the difference in the supplements when dose==2**

```
## [1] "Confidence Interval: -3.79807046333516"
## [2] "Confidence Interval: 3.63807046333515"
## [3] "P-Value: 0.963851588723373"
```

Interpretation: We fail to reject the null hypothesis that there is no statistically significant difference between the effect of the supplements across all dosages. The p-value is greater than 5% which is the level of significance.

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

**Conclusions:**

Orange Juice as a supplement has a stronger effect on tooth length than compared to ascorbic acid when the dosages are 0.5 mg and 1 mg respectively. However, for a higher dosage of 2 mg, there are no statistically significant difference in the effects.

**Assumption:**

(1) The tooth lengths are normally distributed
(2) There are no unmeasured variables affecting the tooth length.