Statistical Inference - Course Project Part 2

Load and visualize the data

Load the data

```
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
```

Summary of the data

```
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
```

Correct variables

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

Exploratory Data analysis

Verify the mean of the len variable by supply method

```
MeanSupp = split(ToothGrowth$len, ToothGrowth$supp)
sapply(MeanSupp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

Create graph

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
      xlab("Supplement type") +ylab("Tooth length")
```

Inferential Statistics

Do the tooth length of the guinea pigs depends on delivery methods? A t test for the difference will be made to test this claim

```
len<-ToothGrowth$len
supp<-ToothGrowth$supp
dose<-ToothGrowth$dose


sapply(MeanSupp, var)
```

```
##       OJ       VC
```

```
## 43.63344 68.32723
```

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

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[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 of this test was 0.06, which is very close to the significance level of 5%. It could be interpreted as a lack of enough evidence to reject the null hypothesis, however it is paramount to account that the 0.05 value of significance is only a convenience value. Furthermore, the confidence interval of the test contains zero (0)

Now we will test the tooth length of the group with vitamin C dosage

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

```
##
##  Two Sample t-test
##
## data:  len[dose == 2] and len[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
```

Now the p-value of this test is 0, a evidence that we can reject the null hypothesis. Therefore we can assume that the means of dosage change from 1mg to 2mg creates an positive effect on tooth length. Furthermore, the confidence interval does not contain zero (0).

After those analysis we can conclude that supplement type has no effect on tooth growth, and increasing the dose level leads to increased tooth growth.