# Statistical Inference Part 2


***Hisam Sabouni***

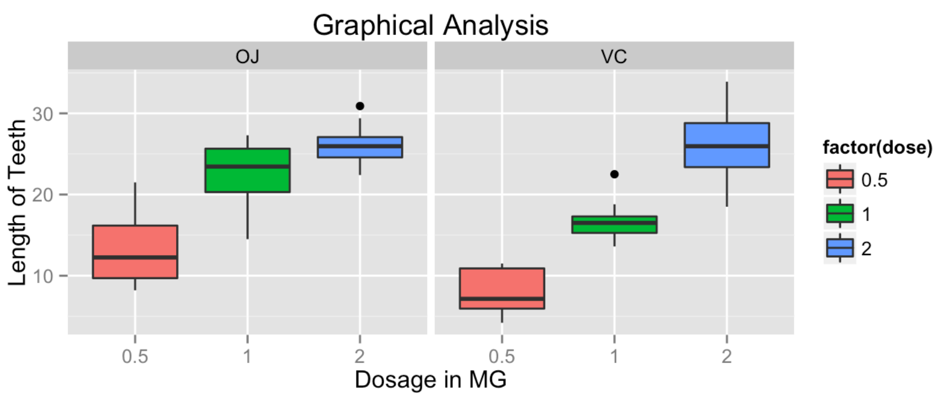

Analysis of ToothGrowth dataset:

#### Load Necessary Packages

```
library(datasets)
library(ggplot2)
```

#### Load the ToothGrowth data and Perfrom Some Basic Exploratory Analyses

```
data(ToothGrowth)
plot.one <- ggplot(ToothGrowth, aes(x=factor(dose),y=len,fill=factor(dose)))
plot.one + geom_boxplot(notch=F) + facet_grid(.~supp) +
    scale_x_discrete("Dosage in MG") +
    scale_y_continuous("Length of Teeth") +
    ggtitle("Graphical Analysis")
```




There appears to be a positive correlation between Length of Teeth and Dosage of OC and VC

#### Basic Summary of Data

```
head(ToothGrowth,n=3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
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 ...
```

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

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

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

We have a total of 60 observations broken apart in to 6 groups of 10. There are two categories, OC and VC, that are seperated out by dosage ranging from 0.5MG to 2MG.

**Statisical Testing**

Supplement Effects: H0: Tooth Growth Supplements Effect = 0 (null hypothesis) H1: Tooth Growth Supplements Effect > 0 (alternative hypothesis) One-sided test with 58 degrees of freedom:

```
supplement.test1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
supplement.test2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
summary.simple.test1<-cbind(supplement.test1[[1]],supplement.test1[[3]],supplement.test1[[4
]][1],supplement.test1[[4]][2])
summary.simple.test2<-cbind(supplement.test2[[1]],supplement.test2[[3]],supplement.test2[[4
]][1],supplement.test2[[4]][2])
summary.simple.test<-rbind(summary.simple.test1,summary.simple.test2)
colnames(summary.simple.test)<-c("T-Stat","P-Value","Confidence Low","Confidence High")
rownames(summary.simple.test)<-c("Equal Variance Test","Unequal Variance Test")
summary.simple.test
```

```
##                         T-Stat    P-Value Confidence Low Confidence High
## Equal Variance Test   1.915268 0.06039337     -0.1670064        7.567006
## Unequal Variance Test 1.915268 0.06063451     -0.1710156        7.571016
```

The T-Statistic, P-Values, and Confidence Interval, all indicate, in equal variance and uneqaul variance tests, that we do not have enough information to reject our null hypothesis at the 95% confidence level. However, we can reject the null hypothesis in favor of the alternative hypothesis at the 90% confidence level; giving us a slightly larger probability of Type-I error (10%).

Dosage Effects: Subsetting the data:

```
dosage.low<-subset(ToothGrowth, ToothGrowth[,3]==0.5)
dosage.medium<-subset(ToothGrowth, ToothGrowth[,3]==1)
dosage.high<-subset(ToothGrowth, ToothGrowth[,3]==2)
```

I will now run three t-tests, 0.5 and 1, 1 and 2, 0.5 and 2, with two cases for each, equal varaince and unequal variance.

Test 1: 0.5MG and 1MG Doses

```
dosage.low.medium.t1<-t.test(dosage.low[,1],dosage.medium[,1],var.equal=F,paired=F)
dosage.low.medium.t2<-t.test(dosage.low[,1],dosage.medium[,1],var.equal=T,paired=F)
dosage.low.medium.1<-cbind(dosage.low.medium.t1[[1]],dosage.low.medium.t1[[3]],dosage.low.m
edium.t1[[4]][1],dosage.low.medium.t1[[4]][2])
dosage.low.medium.2<-cbind(dosage.low.medium.t2[[1]],dosage.low.medium.t2[[3]],dosage.low.m
edium.t2[[4]][1],dosage.low.medium.t2[[4]][2])
dosage.low.medium<-rbind(dosage.low.medium.1,dosage.low.medium.2)
colnames(dosage.low.medium)<-c("T-Stat","P-Value","Confidence Low","Confidence High")
```

```
rownames(dosage.low.medium)<-c("Equal Variance Test","Unequal Variance Test")
dosage.low.medium
```

```
##                          T-Stat      P-Value Confidence Low
## Equal Variance Test   -6.476648 1.268301e-07      -11.98378
## Unequal Variance Test -6.476648 1.266297e-07      -11.98375
##                       Confidence High
## Equal Variance Test         -6.276219
## Unequal Variance Test       -6.276252
```

Test 2: 1MG and 2MG Doses

```
dosage.medium.high.t1<-t.test(dosage.medium[,1],dosage.high[,1],var.equal=F,paired=F)
dosage.medium.high.t2<-t.test(dosage.medium[,1],dosage.high[,1],var.equal=T,paired=F)
dosage.medium.high.1<-cbind(dosage.medium.high.t1[[1]],dosage.medium.high.t1[[3]],dosage.me
dium.high.t1[[4]][1],dosage.medium.high.t1[[4]][2])
dosage.medium.high.2<-cbind(dosage.medium.high.t2[[1]],dosage.medium.high.t2[[3]],dosage.me
dium.high.t2[[4]][1],dosage.medium.high.t2[[4]][2])
dosage.medium.high<-rbind(dosage.medium.high.1,dosage.medium.high.2)
colnames(dosage.medium.high)<-c("T-Stat","P-Value","Confidence Low","Confidence High")
rownames(dosage.medium.high)<-c("Equal Variance Test","Unequal Variance Test")
dosage.medium.high
```

```
##                          T-Stat      P-Value Confidence Low
## Equal Variance Test   -4.900484 1.906430e-05      -8.996481
## Unequal Variance Test -4.900484 1.810829e-05      -8.994387
##                       Confidence High
## Equal Variance Test         -3.733519
## Unequal Variance Test       -3.735613
```

Test 3: 0.5MG and 2MG Doses

```
dosage.low.high.t1<-t.test(dosage.low[,1],dosage.high[,1],var.equal=F,paired=F)
dosage.low.high.t2<-t.test(dosage.low[,1],dosage.high[,1],var.equal=T,paired=F)
dosage.low.high.1<-cbind(dosage.low.high.t1[[1]],dosage.low.high.t1[[3]],dosage.low.high.t1
[[4]][1],dosage.low.high.t1[[4]][2])
dosage.low.high.2<-cbind(dosage.low.high.t2[[1]],dosage.low.high.t2[[3]],dosage.low.high.t2
[[4]][1],dosage.low.high.t2[[4]][2])
dosage.low.high<-rbind(dosage.low.high.1,dosage.low.high.2)
colnames(dosage.low.high)<-c("T-Stat","P-Value","Confidence Low","Confidence High")
rownames(dosage.low.high)<-c("Equal Variance Test","Unequal Variance Test")
dosage.low.high
```

```
##                          T-Stat      P-Value Confidence Low
## Equal Variance Test   -11.79905 4.397525e-14      -18.15617
## Unequal Variance Test -11.79905 2.837553e-14      -18.15352
##                       Confidence High
## Equal Variance Test         -12.83383
## Unequal Variance Test       -12.83648
```

**Conclusion and Assumptions:** From the tests conducted above, one can conclude that higher dosages of either supplement will result increase tooth length. Specifically, 2MG will lead to higher results than 1MG and 1MG will lead to higher results than 0.5MG dosages.